

Syntactic Structure and its Effects  
on Performance in a Second Language

Paul Michael Meara

Thesis submitted for the degree of  
D.Phil at the University of York.

Department of Language

February 1980

**PAGE  
NUMBERING  
AS ORIGINAL**

## Table of Contents

Acknowledgements	1
Abstract	2
CHAPTER 1: Competence, Performance and Research into Second Language Learning	
1-1 On Paradigms in Research in Foreign Language Learning	6
1-2 Two Examples of Competence Research	9
1-2-1 Error Analysis	9
1-2-2 Longitudinal Studies of Individual Learners	13
1-3 Problems with the Competence Paradigm	18
1-4 Studies of Linguistic Performance	25
1-4-1 Some Preliminary Considerations	25
1-4-2 The Effects of Surface Structure on Performance in a First Language	30
1-4-3 Syntactic Structure in a Second Language	35
CHAPTER 2: Experiments with Statistical Approximations	
2-1 Introduction	45
2-2 General Background	46
2-2-1 Short Term Memory and Sentence Processing	48
2-2-2 Memory Span and Foreign Language Learners	52
2-3 Statistical Approximations and Contextual Constraint	55
2-3-1 Statistical Approximations	55
2-3-2 Construction of Statistical Approximations	56
2-3-3 Miller and Selfridge's Experiment	59
2-3-4 Interpretation and Further Work	60
2-4 Statistical Approximations and Foreign Language Learners	62
2-4-1 General Considerations	62
2-4-2 Previous Work with Learners	65
2-5 Experiment 1	71
2-5-1 Materials	72
2-5-2 Subjects	75
2-5-3 Method	75
2-5-4 Results	76
2-5-5 Experiment 1 - Summary	88

2-6	Experiment 2	89
2-6-2	Materials, Method and Subjects	90
2-6-2	Results	90
2-6-3	Summary	92
2-7	Discussion and Comments	95
2-7-1	Coleman Scores	96
2-7-2	Serial Effects in Recall	100
2-8	Conclusions	105
CHAPTER 3: Click Location		
3-1	Introduction	109
3-2	The Click-Placement Paradigm	110
3-2-1	Ladefoged and Broadbent's Experiments	110
3-2-2	The Development of the Click Paradigm	114
3-3	Click Placement and Foreign Language Learners	118
3-3-1	General Considerations	118
3-3-2	Previous Work with Learners	120
3-4	Experiment	123
3-4-1	Materials	124
3-4-2	Subjects	124
3-4-3	Method	125
3-4-4	Scoring	125
3-4-5	Results	126
3-5	Discussion	139
CHAPTER 4: The Need for Structure Specific Tests		
CHAPTER 5: Recognition Thresholds for Phrases		
5-1	Introduction and Background	162
5-2	Previous Studies of Word Recognition in a Second Language	164
5-2-1	Macnamara	168
5-2-2	Morris	174
5-2-3	Summary	177
5-3	Pilot Study	178
5-3-1	Materials	179
5-3-2	Method	187
5-3-3	Results	187

5-4	Main Experiment	192
5-4-1	Materials	193
5-4-2	Subjects	193
5-4-3	Method	194
5-4-4	Results	194
5-4-5	Supplementary Experiment	199
5-5	Discussion	201
CHAPTER 6: Probe Latencies		
6-1	Introduction	212
6-2	Background	214
6-2-1	Suci, Ammon and Gamlin	214
6-2-2	Length Effects	216
6-2-3	Results with Children	217
6-2-4	Meaningless Material	218
6-2-5	Conclusions	218
6-3	Extensions and Developments of the Probe Latency Paradigm	220
6-3-1	Summary	220
6-3-2	Kennedy and Wilkes	221
6-3-3	Ammon	229
6-3-4	Kempen	233
6-4	Probe Latencies and Second Language Speakers	235
6-5	Experiment 1	237
6-5-1	Materials	238
6-5-2	Subjects	240
6-5-3	Method	240
6-5-4	Results	244
6-6	Experiment 2	249
6-6-1	Introduction	249
6-6-2	Materials	249
6-6-3	Subjects	250
6-6-4	Method	251
6-6-5	Results	254
6-7	Discussion	256
6-7-1	Some Possible Explanations	256
6-7-2	Two Further Possibilities	260

CHAPTER 7: Conclusions.

7-1	Summary	269
7-2	Positive Findings	274
7-3	The Centrality of Syntax: a Reinterpretation.	280
7-4	Universal Psycholinguistics?	291
7-5	Suggestions for further work	302

APPENDICES 306

2-A	Materials used in Chapter 2	307
2-B	Scores of Individual Subjects	309
3-A	Materials used in Chapter 3	311
3-B	Percentage of clicks correctly located	313
3-C	Basic Displacement scores	314
3-D	Corrected displacement scores	315
5-A	Materials used in pilot study	316
5-B	Scores of individual subjects in pilot study	317
5-C	Materials used in main experiment	318
5-D	Scores of individual subjects in main experiment	319
6-A-1	Materials used in Experiment 1	320
6-A-2	Scores of individual subjects in Experiment 1	322
6-B-1	Materials used in Experiment 2	323
6-B-2	Scores of individual subjects in Experiment 2	325
6-B-3	Error rates. Experiment 2	326
7-A	The Language of the experiments	328
7-B	The level of proficiency of the subjects	336
7-C	Foreign Language Learners and other Non-standard Populations.	347

FOOTNOTES 356

BIBLIOGRAPHY 363

## ACKNOWLEDGEMENTS

I should like to thank those whose support and encouragement made this work possible:

Dr. D. Reibel and Dr. S. Harlow, who supervised the work in its initial and final stages respectively;

Dr. D. Bruce for his invaluable and unstinting advice and comments;

my colleague Dr. M. H. A. Blanc for considerable moral and material support;

A. S. Beevor who typed the manuscript;

and Dr. G. M. Awbery, whose influence will be apparent throughout the work.

I should also like to thank Mr. J. L. M. Trim who first taught me linguistics and persuaded me of the importance of research into second language learning.

## ABSTRACT

The central argument of this thesis is that the greater part of research in Applied Linguistics is concerned with questions which are based on notions of linguistic competence, and that a number of important areas concerned with performance in a second language have been unjustly neglected. The shortcomings of the Competence approach are discussed, and a number of well-known phenomena which cannot easily be handled by such an approach are considered. A case for detailed study of aspects of performance in a second language is then made.

Four studies of second language performance are then reported. These all concern the effects of surface structure syntax on the behaviour of second language learners, and each study compares groups of second language learners with native speakers. The experiments make use of a number of experimental paradigms: immediate recall of statistical approximations, click location, tachistoscopic recognition of phrases and a probe-latency task. All the tests show marked differences between native speaker performance and the performance of the learner groups. The implications of these findings are discussed.

The results are used to assess the claim made by Macnamara and by Trim that foreign language learners fail to make proper use of syntactic redundancies when they are operating in a second language. The results partly support this claim, but not entirely.



A number of major problems emerge with the data, which suggest that the questions asked may to some extent be premature. In particular, a case is made for the view that some widely accepted claims, put forward on the basis of experiments conducted on English and closely related languages, may not hold for other languages. The data also point to word-recognition as a major source of difficulty for second language learners.

Doing psycholinguistic research is hard. All but the gross features of linguistic theories are vague and in dispute; and in psychology there is an unfortunate tendency to substitute disputes about how to get and evaluate results for disputes about what the results imply. In consequence, the research that is considered sound (if it relates to the psychology of language at all) merely demonstrates that the obvious aspects of linguistic theories have behavioural manifestations, and the research about theoretically more risky issues tends to be unbelievably crude.

Wales, and Walker. (1977)

CHAPTER 1.       COMPETENCE, PERFORMANCE, AND RESEARCH INTO  
                  SECOND LANGUAGE LEARNING.

1-1   Paradigms in Research in Foreign Language Learning.

1-2   Two examples of Competence Research.

1-3   Problems with the Competence Paradigm.

1-4   Studies of Linguistic Performance.

1-1. ON PARADIGMS IN RESEARCH IN FOREIGN LANGUAGE LEARNING.

A major difficulty to be faced in undertaking theoretical research into second language learning is that many eminent applied linguists have seen themselves not as people concerned with important questions of theory, but rather as mere passers on of insights from theoretical study into the nature and structure of language. Corder (1973), for example, explicitly asserts that Applied Linguistics "is not a theoretical study" (p10) and he goes on to reinforce this claim by elaborating a rigidly hierarchical model of information flow in which insights and ideas generated by theoretical linguists are transmitted by the applied linguist to teachers who in their turn facilitate the task of actual learners by implementing these insights in practical situations.

Not surprisingly, then, given that this attitude is a fairly wide-spread one, where theoretical work has been undertaken, it has been heavily dependent on constructs taken over from theoretical linguistics. Particularly influential in this respect has been Chomsky's distinction between Linguistic Competence and Linguistic Performance (Chomsky 1965 and elsewhere).

The importance of the competence/performance distinction for theoretical linguistics is that the distinction is a simplifying assumption which makes it possible to treat some limited aspects of language behaviour in a coherent way, while allowing us to ignore other problems which are less immediately tractable. Thus, theoretical linguists typically ignore hesitation phenomena, false starts and other disfluencies and ungrammatical sequences produced

in the full flow of speech, together with "memory limitations, distractions, shifts of attention and interest and errors (random or characteristic)" (Chomsky 1965 p3) which are considered to lie outside the scope of the study of linguistic competence. Simplifying assumptions of this sort are fully justified in so far as they lead to significant theoretical advances, and as far as theoretical linguistics is concerned, there can be little doubt that significant advances have arisen out of the study of linguistic competence - the criticisms of scholars such as Derwing (1973) notwithstanding.<sup>1</sup> In the study of foreign language learning, however, it is less clear that this framework serves any useful purpose. Rather it seems that our over-dependence on the notion of linguistic competence has led to the development of a research paradigm (in the sense of Kuhn 1962), in which some important questions which do not fit readily into the conceptual framework imposed by the paradigm have been totally ignored, while other secondary problems, which do fit this same conceptual framework rather more closely, have been given a prominence that is not wholly deserved.

The great bulk of what one might call theoretical applied linguistics - that is research into the theoretical questions concerning the learning of second languages - has been carried out within a framework of this kind, where the central concern is the study of the linguistic competence of the learner. Let us call it the Competence Paradigm. The sort of question that has been seen as important within this paradigm relates primarily to the development of the learner's competence in his foreign language, rather than to any other aspect of his linguistic behaviour. So, typical

questions of this sort would include:

- . What stages does the learner's developing competence pass through?
- . Is there a natural order for the development of syntactic structures and morphemes in naturalistic language learning?
- . If there is a natural order, how closely does it correspond to what we know about the development of these structures in first language acquisition?
- . What light can be thrown on the development of learner's competence by a study of the errors that he makes?
- . Is it possible to predict the errors that learners will make by comparing the grammars of the target and native language?

and so forth.

This type of approach is epitomised by two types of study in particular, which between them account for the greater part of recent research in Applied Linguistics, and are widely considered to be the mainstream of Applied Linguistic research. Quantitatively, the more important of these are studies of errors - cf Valdman 1975 for a review of this field. The second field, quantitatively smaller, but now becoming a major growth area, concerns longitudinal studies of the development of second languages in individual subjects.

The next two sections will discuss each of these types of study, and show how their concerns are dominated by notions drawn from the competence paradigm. The final section of this part will then discuss some problems which cannot be easily handled by the competence paradigm, and a rather different type of approach based on wider performance factors will be advocated.

## 1-2. TWO EXAMPLES OF COMPETENCE RESEARCH.

### 1-2-1 Error Analysis.

Although not the earliest statement of the principles and methods of error analysis, Corder (1967) is one of the clearest and certainly one of the most influential papers in this field.

Corder pointed out that the errors made by learners had at that time been accorded only scant attention, and had been looked upon mainly as a nuisance, probably avoidable, but in any case to be eradicated as quickly as possible. He suggests that this attitude might be oversimple, and argues that far from being random and of little wider interest, learners' errors can actually be a rich source of data, both for the researcher and the teacher.

Corder is quite explicit about situating his notion of error firmly within the Competence Paradigm. He draws an important methodological distinction between true errors and mistakes, describing the latter as "adventitious artefacts of linguistic performance" due to "memory lapses, physical states such as tiredness and psychological conditions such as strong emotion." As such, mistakes "do not reflect a defect in our knowledge of our language."

Corder goes on to assert:

"We must therefore make a distinction between those errors which are the product of such chance circumstances as these, and those which reveal (the learner's) underlying knowledge of the language to date, or as we may call it, his transitional competence. The errors of performance will characteristically be unsystematic, and the errors of competence systematic..... It will be useful hereafter to refer to errors of performance as mistakes while reserving the term error to refer to the systematic errors of the learner from which we are able to reconstruct his knowledge of the language to date, i.e. his transitional competence." (p24).

The language here is unmistakably that of Chomsky, and the defining criteria of mistakes echo very closely the form of words used by Chomsky in his discussion of the characteristic features of linguistic performance in "Aspects". It is clear that Corder believes that it might be possible to write generative rules which could account for the errors produced by a given learner, and that the differences between these rules and the rules needed to describe the competence of a normal native speaker could be an important source of insights into the nature of foreign language learning. Corder does not give any account of how one should set about subjecting errors to the "systematic analysis" necessary for reconstructing the learner's transitional competence, nor does he discuss any of the problems involved in making inferences about competence on the basis of performance data - again, a failing which is characteristic of work in this paradigm. In a later paper, however, (Corder 1973), he gives the following example of how an error analysis might proceed:

"The learner who said 'I am here since three o'clock' did not just select the wrong grammatical item, but showed that he had not learned the function of the auxiliary system in English e.g. the rule:

Auxiliary → Perfect + tense + since + point of time  
i.e., the verb is in the perfective form when it is  
collocated with the prepositional phrase of time  
since + point of time." (p279)

This example is fairly typical of the way actual errors are handled by error analysis, and though other writers are more rigorous in their approach, their basic preoccupations are identical. The influence of the Competence Paradigm is apparent here in three facets of Corder's approach. Firstly, it emerges in the choice of what data is to be studied. In any error analysis, the raw data is



the sentences produced by a learner or a group of learners, rather than any other aspect of their behaviour in the foreign language such as their ability to comprehend spoken language, or their ability to engage in social interactions in the foreign language. Transformational Grammar also takes sentences as its raw material, and ignores other types of language behaviour in native speakers, so that what we have here is an instance of the concerns of theoretical linguistics being transferred to applied linguistics. To someone not familiar with this paradigm, however, the aim of writing a grammar to account for the sentence forms produced by learners might seem to be an unnecessarily restricted goal, since it ignores a number of aspects of the learner's total range of behaviour in the foreign language which are at least of equal importance. The second way in which the influence of the Competence Paradigm is apparent lies in the transformational formulation of the rule that accounts for the learner's error. This is not strictly necessary, but the majority of error analysts seem to prefer to use formalisations of this sort. (Corder's rule is in fact incorrectly formalized, and the rule as it is stated does not correspond to the ordinary language statement of the rule that follows it.) The third and most basic sign of the influence of the Competence Paradigm is, of course, the way that Corder explains the error in terms of rules at all. He first argues that there is a grammatical rule that is part of the native speaker's competence, and claims that this rule has not been applied when the error in question was produced. It is then argued that the learner's transitional competence in English does not contain this rule, and some more sophisticated error analysts produce alternative rules which can account for the form of the erroneous sentence. There

are a number of difficulties with this sort of explanation, notably problems associated with what criteria one uses to choose between the many possible sets of rules that could describe the native speaker's competence, or the many possible descriptions of the learner's error. It is significant, for example, that Corder's rule is not one that would find its way into any linguists' grammar of English. Rather it is an over-simplification from a much larger set of more complex data, which many linguists would want to take issue with. However, these problems will not concern us here, rather our intention is merely to draw attention to the very close links between research of this type, and the concerns of the dominant paradigm in linguistics from which it draws its theoretical concepts and its method of working. As we shall see later, there are a number of important and interesting problems that a rule-based, competence-orientated approach of this sort cannot handle at all easily.

## 1-2-2 Longitudinal Studies of Individual Learners.

The second type of study in which the influence of the Competence Paradigm is readily apparent is again one that was mooted in Corder's (1967) paper - longitudinal studies of individual learners. In this paper, Corder was already drawing comparisons between the work of people studying the development of language in young children, and those working on foreign language acquisition, suggesting that some of the problems might be similar, and that much could be learned by using research techniques developed for the study of first language acquisition and applying them to second language learning situations. In a later paper (Corder (1971)) Corder developed this line of thinking, and argued that classical error analysis of the type discussed above was essentially limited in that it could provide only static, synchronic pictures of the learner's transitional competence, and gave no insight into any of the diachronic variations in transitional competence - how it varied with continued exposure to the target language. Corder suggested that a way round this problem might be to study individual learners very closely over long periods of time, and that their foreign language output should be monitored and analysed using methods which were at that time being widely used in studies of first language acquisition. These methods were of two main types. Some first language studies (e.g. Bayes (1969)) had attempted to monitor the whole of a child's output over long periods of time, clearly a task of mammoth proportions, which requires no small amount of organization and control over the child's environment. With adults, the difficulties of monitoring language output in this way are almost insuperable, and so not surprisingly studies of this kind in a second language are very rare. One good example of work

of this type, however, is Hughes (1979) who studied the entire English output of a single native Spanish speaker over a period of some nine months. This feat was achieved only at the cost of imposing severe limitations on the occasions when the learner was allowed to use English, a constraint which seems artificial and unnatural, but which in any case must severely restrict the generality of the findings.

A more workable, and therefore more common approach to the acquisition of a first language consists in monitoring the child's output at regular intervals, and studying the development of selected parts of the child's grammatical competence. Brown's (1974) study of three children is probably the best known example of this type of research. McNeill (1970) describes the approach as follows:

"Rather than attempt to describe the total corpus collected from a child at some point of time, one examines the emergence of a particular grammatical system as it is manifested at different times.....(this) strategy starts with a part of the adult grammar, and judges if there is sufficient evidence in the corpus to justify ascribing it to the child." (p9).

This technique is readily adaptable to work with foreign language learners, especially to the study of learners in naturalistic settings, where the artificial constraints of the classroom and the preplanned syllabus are not relevant conditions. Studies of this type are now becoming increasingly numerous, and much of this work is explicitly concerned with testing the claim that natural developmental orders for the emergence of morphological features run parallel in children's acquisition of English as a first language and adult acquisition of English as a second language. (cf. for instance Dulay and Burt (1974) or the review of this whole field by Corder (1978).)

An early, but classic illustration of this type of approach is the work of Dato (e.g. 1970 and elsewhere). In this study, five young American children living in Madrid were observed. Each of the children was learning Spanish, but no formal instruction was provided. They played frequently with other children who were native speakers of Spanish, and on the strength of this exposure, were expected to pick up some knowledge of the language. Dato collected regular samples of the children's speech in a series of "semi-structured sessions" lasting about 30 minutes every two weeks, and the transcripts of these sessions were then analysed along the lines suggested by McNeill. Dato's (1969) paper summarizes the development of the Auxiliary construction in his children, and this paper will be discussed here as a further illustration of the extent to which constructs absorbed from the Competence Paradigm have influenced the sort of questions that are considered worthy of research in Applied Linguistics. The paper takes as its starting point a transformational analysis of the Verb Phrase in Spanish. The status of this description is not wholly clear, as it is an amalgam of the work of a number of theoretical linguists, rather than a carefully argued and coherent description, but alternative analyses are not discussed. Dato then goes through the records of his children's speech, and discusses the order in which the various features of the theoretical model appear in the transcripts. He claims that the children's Auxiliary rules gradually become more and more complex, as they approximate more closely to the complete adult rule, and he further argues that all the children in his sample, irrespective of their age, pass through a very similar developmental sequence. He uses this finding to suggest that there might be a natural order of development for the features of the

auxiliary structure, an order which is arrived at by establishing the order of development in each of the four children studied, and calculating a mean rank order of emergence for each feature on the basis of this data. In a series of later papers, Dato extends his analysis to cover the development of the Noun Phrase, Embeddings of various types, and so forth. The methods used and the substantial points made are basically the same, however.

There are a number of serious problems with this work - notably the difficulty of establishing a reliable order of emergence for syntactic features when only four subjects are studied (no statistical tests are invoked to assess the final order put forward by Dato,<sup>2</sup>) and when the status of the features studied is itself unclear. These problems will not concern us here, however. What does emerge clearly, even from this very brief outline of Dato's work is the pervasive influence of the Competence Paradigm. Just as we saw above in the case of Error Analysis, the paradigm shows itself in the decision to take sentences as the basic data source, in the decision to account for the sentences by writing increasingly complex sets of rules which generate the sentence forms, and in the formulation of these rules in transformational terminology. Furthermore, Dato's own discussion of his work, is quite explicit about using this competence-orientated framework. He begins his discussion by claiming that his paper has "concentrated on the description of the actual process by which certain children have learned Spanish" (p21, emphasis added), a claim which, given the nature of the data studied and the sort of analysis applied to it, clearly equates learning a foreign language with acquiring the necessary rules of competence, and implies that once these rules of competence have been acquired little remains to be accounted for.

Later, Dato asserts that:

"one of the most insightful observations that can be made in language acquisition is to relate children's actual utterances to the intermediate structures generated as a result of applying certain base and transformational rules as described in generative grammar" (p23).

Why such observations should be considered insightful is not made clear, but it should be obvious that a statement of this type makes no sense outside a framework which is not centrally directed towards the study of linguistic competence.

### 1-3. PROBLEMS WITH THE COMPETENCE PARADIGM.

In the two preceding sections, we discussed the main strands of theoretical research in Applied Linguistics, and it was argued that this research is heavily dominated by what we have referred to as the Competence Paradigm.

One of the principal drawbacks of paradigm research, according to Kuhn (1962) is that a paradigm forces attention to be focussed narrowly onto the kind of question which could in principle be answered by the constructs that the paradigm makes available. The Competence Paradigm is mainly concerned with questions of well-formedness and the rules which generate sentences, and this leads to research in Applied Linguistics being directed towards questions which can be easily answered using concepts such as rules and rule-ordering, and intermediate stages in transitional competence. The disadvantage of this is that certain other questions which are not amenable to study in terms of rules and competence constructs, and do not fall within the scope of the paradigm, are typically ignored by the paradigm's adherents, and treated as questions of little importance or interest. The paradigm thus defines a research area, and provides tools for handling these questions, but in so doing, it invalidates a whole set of other questions.

This narrow funnelling effect of the paradigm is very easy to observe in Applied Linguistic research. The channelling of research effort into questions of Competence, such as those outlined above, has led to a large number of questions being ignored, despite the fact that many of these questions are in some ways more obviously



relevant to the practical issues involved in teaching and learning languages than the problems that currently lie in the mainstream of research.

A very clear example of this sort, a problem which has failed to attract the attention that it deserves, is the area of Vocabulary Acquisition in a Foreign Language. Once they are beyond the most elementary stages of learning a language, almost all learners identify this area as their greatest single source of difficulty. This in itself ought to be sufficient grounds for much serious work on the problems of vocabulary acquisition to have been undertaken. In fact, however, very little research of any substance is to be found, and even this work often makes assumptions which are naïve and misleading, such as Raugh and Atkinson's implicit assumption that learning vocabulary in a foreign language can be adequately described in terms of a paired associate model of learning taken over uncritically from studies of verbal learning in psychology. (Raugh and Atkinson 1975. For a bibliography of recent research in vocabulary Research, see Twomey 1979).

Transformational grammarians typically have little to say about the lexicon, and even less to say about the development of the lexicon in individual speakers. Recent developments in theoretical linguistics have left this area largely unexamined, and Applied Linguistics, with its dependence on this Competence Paradigm have likewise not considered the important issues that questions of vocabulary acquisition raise. Indeed text-book writers seem to feel that they can ignore this question with impunity, and it is rare for vocabulary acquisition to be given more than passing consideration. Wilkins (1972) for instance exceptionally devotes an entire chapter

to vocabulary questions, but the greater part of this discussion is given over to consideration of frequency counts as a basis for choosing the vocabulary content of courses and tests. The nature and structure of the learner's second language lexicon, and how it is accessed are not discussed at all. A more extreme but in some ways more typical instance of the trend, is Bennett (1974), a book primarily intended for teachers, and very heavily indebted to the Competence Paradigm. In all its 350 pages, this book contains only one reference to vocabulary, quoted here in full:

"Vocabulary extension exercises would exploit the derivational system of language which native speakers KNOW and use, in some cases particularly adventurously." (p303 sic.)

This comment hardly does justice to the importance of vocabulary at the theoretical or the practical level, but does clearly indicate how easily a problem of this importance can become the victim of a paradigm.

Vocabulary acquisition is not, of course, the only area to have suffered in this way.. There is, in fact, a whole range of phenomena of this sort, which seem to be inherently incapable of explanation on the terms provided by the Competence Paradigm. We know, for example, that learners typically experience considerable difficulty in handling long sentences, and that they also have some difficulty in understanding sentences spoken at speed. Native speakers do not seem to experience these difficulties to anything like the same extent, and this discrepancy suggests that learners might be operating under some sorts of constraint that do not normally impinge on the performance of native speakers. The difficulty with long sentences suggests that some form of limitation on short term memory capacity is involved, or some restriction in the capacity of any one

or more of the many intermediate stores that have been proposed as part of the sentence processing mechanism. Presumably, normal intercourse is so structured that normal native speakers' memory capacity is never over-loaded, but it is not difficult to see that if foreign language learners had reduced memory capacity, they would frequently be faced with sentences which exceeded this capacity, and resulted in a failure of understanding. Long sentences would be obvious candidates for this class, and so also might be some sentences which are complex and required large amounts of computing space for reasons other than sheer length. Memory limitations of this type are explicitly discounted by the Competence Paradigm, since Chomsky lists this as one of his "grammatically irrelevant features" which are considered to be a part of performance.

Again, the difficulty with rapidly spoken sentences suggests that the rate of flow of information is an important consideration for learners. It seems plausible to suggest that the rate of information flow in sentences for a population of native speakers varies between both an upper and a lower limit. Sentences spoken so slowly that the mean information rate lies outside the normal lower limit would fail to be understood because it would just be too difficult to keep one's attention focussed. Sentences spoken so quickly that the upper bound was exceeded would fail to be understood if the processing capacity of the speech understanding mechanisms proved to be insufficient. Normal speech presumably accommodates itself to these limitations, in such a way that they are unconsciously absorbed, and become a distinguishing characteristic of our speech behaviour, without our being aware of them. The fact that learners cannot cope easily with fast speech suggests that their range of tolerance for information flow is much more narrow

than the range of tolerance that we assume to operate for native speakers.

It is extremely difficult to see how a rule-based approach to the problems of language learning, which takes its main theoretical constructs from the Competence Paradigm could make any meaningful statements about phenomena such as these.

Despite the fact that difficulties of this sort are commonly experienced by any foreign language learner, and despite the existence and easy availability of experimental techniques which could be readily used for investigating the phenomena, there has been no systematic attempt to study the causes or the limits of these difficulties, or how their effect could be reduced or circumscribed, and how, in consequence, the performance characteristics of the learner improved. There is, in fact, a general assumption that all problems of this type are problems of competence, in that failure to handle sentences correctly (either in production or in reception) is taken to mean that the learner does not know the structures involved, and that his competence is deficient in this respect. Underlying this account is an assumption that Competence is an all or nothing affair, that the learner either knows the structure in question or not. It is easy to show, however, that this assumption is something of an over over-simplification. It is commonly known, for instance, that for most learners receptive skills are usually more advanced than productive skills, and this suggests that a competence that is well-developed in one modality may be undeveloped and inadequate in another modality. Thus a learner might be able to read, but not to produce the sentences that he is able to read in a spoken interaction. To suggest that there are several independent competences all

developing simultaneously and possibly along dissimilar lines, seems to be an unparsimonious way of accounting for these facts, in that it multiplies entities in an uneconomical way. A simpler explanation probably is to be found in the nature of the tasks, which make different types of demands on the learner, and thus allow his competence to be used more or less effectively. However, considerations of this type point us away from Competence, and towards a consideration of the types of performance constraints that impinge on behaviour in a foreign language.

The same problem arises even more acutely with advanced foreign language learners. Consider, for example, the case of the "Fairly Fluent Learner", a learner whose competence is in most respects indistinguishable from that of the native speaker, familiar with the syntax of the language he is learning, and with a vocabulary large enough that he is rarely stuck for a word. Such a learner is unlikely to experience any difficulty in reading general texts in his foreign language, and he may probably be able to write and speak fluently too. Yet there are certain predictable situations in which learners of this sort typically fail to perform like native speakers. We know, for instance, that fluent learners experience considerable difficulty in handling foreign language material if it is presented in noisy conditions (Spolsky et al 1968) and that relatively low noise levels which have a negligible effect on native speakers seriously effect the fluent learner's ability to understand the spoken word. Similar difficulties are experienced in Cloze Tests, where words are deleted from a written passage and blanks substituted, and the task requires a suitable word to be supplied for each blank. Oller (1972) has shown that tests of this sort reliably discriminate between native speakers and advanced learners of a

language, even where other tests fail to do so. In general, it seems to be the case that wherever learners are expected to handle degraded stimuli, or in situations where only limited processing time is available similar difficulties arise. Again, it is difficult to see how a competence based approach could handle difficulties of this sort. If the Fairly Fluent Learner can read adequately and write fluently, then his competence must in some respects be fully developed. The problem of why this competence apparently fails to function adequately in difficult circumstances is again one that cannot be resolved within a framework that does not consider performance factors.

#### 1-4. STUDIES OF LINGUISTIC PERFORMANCE

##### 1-4-1 Some preliminary considerations

So far we have discussed the two main types of study that dominate theoretical work in Applied Linguistics. We have argued that this work draws its main theoretical constructs and its principal concerns from the Competence Paradigm that has been taken over from Theoretical Linguistics, and we have further argued that there are a number of phenomena which do not appear to be readily explainable in terms of competence constructs, and seem to require an explanation based on Performance factors.

The remainder of this thesis will be concerned with an investigation of one major aspect of linguistic performance: viz. whether the surface structure syntax of sentences and phrases has the same effects on the behaviour of second language speakers as it does on the behaviour of speakers in their native language.

Some justification for this choice is clearly needed at this stage, since as we have seen, the Competence Paradigm excludes a large number of different and interesting areas, all worth studying. The main reason for concentrating on the behavioural effects of syntactic structure, rather than any of the other many possible aspects of performance which could form a suitable area for study, is that this area is one which has been the object of considerable attention in recent years. The basic phenomena with respect to native speakers, at least, are well-established, even if they are not well-understood. This seemed to be an important consideration and a major advantage compared to other potentially fruitful fields such as the study of hesitation phenomena, where the data is fragmentary,

and where there is dispute over such fundamental concepts as how long a silence must be before it is considered to constitute a pause. (cf Dechert 1979).

The second reason for choosing to study the behavioural effects of syntactic structure was that this seemed to be a natural extension and development of the Competence Paradigm into the Performance field, rather than a complete break with the paradigm. The work makes use of the constructs of the Competence Paradigm, but at the same time it stands aside from the Paradigm, and critically assesses these constructs, rather than accepting them as given. The "reality" of these constructs or their usefulness as explanatory principles are not taken for granted. Rather, the question is asked whether these constructs can be shown to have any meaning in the wider context outside the descriptions produced by linguists, and an answer to this question is sought by trying to show that the constructs have measurable effects on the linguistic performance of real people.

Each of the three major components of the transformational description of sentences have already been studied in this way: Surface Structure, Deep Structure, and the Transformational Rules by which these two levels of description are mapped onto each other. The work reported in the following sections is chiefly concerned with the first of these components. Deep Structures and Transformations are almost completely ignored. There are two reasons for focussing attention on Surface Structure rather than on either of the other two components.

Firstly, Surface Structures are, in a sense, logically prior to Deep Structures, in that they involve less abstraction and are more



closely related to observable phenomena than Deep Structures or Transformations, which can only be arrived at by a complex process of inference. These latter constructs do not make much sense without the idea of surface structure, and there does not seem to be much point in studying the role of higher order abstractions in linguistic performance, unless you have good grounds for believing in advance that the Surface Structure description of sentences has some kind of role to play in performance too. As we shall see, Surface Structure syntax can be shown to have measurable effects on the behaviour of native speakers of a language, but it is an empirical question whether the same effects are to be found in non-native speakers. What little evidence there is available on this point, is inconclusive. It would in any case be necessary to make an assumption that surface structure was having similar effects in native speakers and learners of language in order to establish any sensible interpretation of a similar Deep Structure or Transformational effect in the performance of learners and native speakers, and it therefore seemed prudent to begin by examining the viability of this assumption, rather than by investigating the effects of Deep Structures and Transformations in the first instance.

The second reason for choosing to study the effects of Surface Structure syntax rather than Deep Structures or Transformations, is that there is considerable disagreement among linguists as to what the "Correct" Deep Structure of any sentence is, and even greater disagreement over the Transformational rules that relate these two forms to each other. This means that psycholinguistic work in this area must choose either to follow a particular school of linguistic thought without critically assessing the descriptions

it uses, or to limit itself to the study of syntactic features which are not in dispute. The first course of action leaves one open to the risk of being bypassed by new developments in the theory. Much of the psycholinguistic work of the 1960's that followed from George Miller's (1962) paper - enormously influential in its time - has now been discredited in this way, because it used models and conceptualizations which are no longer considered viable by theoretical linguists. The alternative course of action, to concentrate on non-controversial features of syntax, is much more cautious, and tends to produce results which are of only limited interest. Much of current psycholinguistics falls into this category. A Good example of a non-controversial issue of this kind is the suggestion that Deep Structure Clauses function as units of perception. (cf. Clark and Clark 1977). The problem here is that clauses are a relatively high level of organization, and it is in no way surprising that very large units of this sort should function as wholes in behaviour. The very large amount of work being done at this level ignores the potentially much more interesting fine detail of Deep Structure Processing, however, leaving unanswered such questions as how clauses are identified from their constituent parts, and their smaller constituents built up into manageable wholes.

These two problems taken together, then, seem to provide sufficient justification for concentrating exclusively on Surface Structure Syntax in what follows. This decision should not be taken as implying that there is no disagreement among linguists over how to describe the surface structures of sentences, however. On the contrary, the criteria for producing a description of the syntactic structure of sentences are not well-defined, and it is possible to find reasonable alternative syntactic descriptions for the same sentence

which differ quite widely, even though they are based on the same structural assumptions about syntax. In practice, however, there is not a great deal of disagreement among linguists over this matter, though this is due more to an overriding theoretical concern with Deep Structures and Transformations rather than any real understanding of surface structures per se. Rather surprisingly no serious discussion of the way Surface Structure descriptions are assigned seems to have taken place. This deficit, coupled with the tendency of psychologists to treat variables such as syntactic structure in a somewhat cavalier and uncritical way means that claims about the effects of syntax on language behaviour cannot always be taken at face value, and it is important that this caveat should be borne in mind in the discussion that follows.

#### 1-4-2 The effects of Surface Structure on Performance in a First Language

Evidence for the claim that the surface structure constituents of sentences and phrases have some sort of psychological reality comes from two principal sources.

Firstly, there exists a large body of evidence which demonstrates that there is a distinction to be drawn between unstructured lists of items and sequences of items which can be seen as syntactically structured. Though this distinction is clearly a very coarse one, the findings on which it is based are well-established and robust, and are drawn from a large range of experimental paradigms. Thus, Epstein (1961 and 1962) found that nonsense words with grammatical inflections added to them in such a way as to produce pseudo-sentences in English are considerably easier to learn than the same sequences of nonsense words without the grammatical endings. Glanzer (1962) similarly found that pairs of nonsense words linked by English function words so as to form recognizable constituents were more easily learned than the same pairs of nonsense words without the grammatical function words. With real English material, memory span for sentence material considerably exceeds memory span for unrelated lists of items (Reed (1924) and Brener (1940)), and recall of sentences is qualitatively different from recall of unrelated word lists, in that the serial position effects associated with word lists are generally absent from Subjects' attempts to recall sentences (Deese and Kaufman (1957)), and the clustering effects found with lists of semantically related words are also absent from recall of syntactically structured material. Eye-voice span is greater for sentences than for lists of words (Lawson (1961)),

Morton (1964)) and shadowing sentences is much more accurate than shadowing word lists or scrambled material (Miller and Isard (1963)). Some other examples of structural effects of this type will be discussed in chapters that follow. A large number of other examples can be found in Oldfield and Marshall (1968).

These experiments all show clearly that sentences and phrases are handled rather differently from unstructured word lists by normal adult native speakers of a language. The rather blunt nature of the experimental tools used in this type of research makes it difficult to be precise about the exact nature of the difference, and a number of conflicting models have been proposed to account for the facilitating effects of syntactic structure on performance. Some of these ideas will be discussed in more detail in subsequent chapters. About the central finding that syntactic structure generally facilitates performance there is little dispute, however.

The second source of evidence for the claim that surface structure syntax of sentences and phrases has demonstrable effects in linguistic behaviour is a very large number of experiments in which performance on different sentence types is compared. Typically, experiments of this kind produce results which vary from one sentence type to another, with a close correspondence between the differences in the results and differences between the syntactic descriptions of the test sentence types.

This literature is large and impressive, covering a wide range of experimental tasks and stimulus material. Johnson (1965) for example, showed that errors made in learning lists of sentences appear to be sensitive to constituent structure, in that if one word is correctly recalled, the likelihood of making an error on the next word was considerably higher when the next word belonged to a

different constituent than when it belonged to the same constituent as the preceding word. Levelt (1970) found that subjects asked to rate the degree of closeness between pairs of words in a sentence scale these relationships in a way which closely mirrors the canonical syntactic description of the sentence - and a similar result was reported by Martin (1970) on a sorting task. Hakes and Foss (1970) showed that response latencies in a phoneme monitoring task were effected by the syntactic structure of the string in which the phoneme was embedded. Kornfold (1973) used a probe task, presenting her subjects with a sentence followed by a probe word, and asking them to state whether the probe word did or did not appear in the test sentences. Latencies of correct responses again were heavily influenced by the syntactic structure of the test sentences. Other studies of this type will be discussed in detail in subsequent chapters. General overviews of this work will be found in Levelt (1974), Fodor, Bever and Garrett (1975) and Horton and Jenkins (1971). In general, these experiments represent an important advance on the studies discussed in the previous section. They demonstrate not just that syntactic structure seen as a global phenomenon has measurable effects on performance, but that the details of the syntactic descriptions of sentences and phrases can also be shown to affect subjects' performance on language tasks provided that sufficiently sensitive measuring tools are used.

An important caveat which needs to be made here is that these claims hold only for normal adult native speakers of English, and it should not be assumed that other groups will perform in a similar fashion. There is, for example, a large body of evidence suggesting that non-normal adults are much less responsive to syntactic structuring than would be expected of normal subjects. The best illustration of this

is the considerable amount of work that has been done on the language behaviour of schizophrenics - cf for instance, Pavy (1968), or Maher (1972) who suggests that a general insensitivity to syntactic structure is a typical linguistic symptom of this type of disorder. Certain other types of mentally abnormal adults also show similar characteristics (Levy and Maxwell (1968)) and so too do those otherwise normal speakers who have begun to become aged or senile (Craik and Massani (1967)). At the other end of the scale, young children do not show a clear syntactic effect in some experimental paradigms. For instance, they fail to show improved scores on recall of statistical approximations to English (Bruce and Pugh (1966)) (see chapter VII for a further discussion of this work). It has also been suggested by Hayhurst (1968) that the well-known shift from word association responses which are predominantly syntagmatic to ones which are predominantly paradigmatic may be a phenomenon which is related to children's increasing awareness of word classes and syntactic structuring. (cf Ervin Tripp (1961)).

Two other parts of our caveat, that the results hold (a) for English and (b) for native speakers of that language, are more problematical. Practically the whole of the research work carried out in this area has made use of stimuli in English, and experience with other languages is limited. The only large-scale exception to this claim is the work of Levelt (e.g. Levelt 1974) which uses Dutch stimuli. This work represents only a fraction of the total amount of research work carried out in this area, however. Furthermore, it should be pointed out that Dutch is in many respects closely related to English, and it is often possible to produce English equivalents for Dutch sentences which match not just on a word to word basis, but even produce matches for each morpheme. Very many of the examples

discussed by Levelt (1974) are of this type. Rather surprisingly, research of the type we have discussed, in languages which differ widely from English, is practically non-existent. This is a problem to which we will have reason to return later. For the moment, however, we assume that results which hold for English will probably hold for other languages too, at least insofar as these other languages behave like English behaves.

The fourth part of our caveat is the principal question with which this thesis is concerned - viz. whether non-native speakers of a language are sensitive to its syntax in much the same way as native speakers are, or whether the understanding of phrases and sentences in a foreign language is an activity that is markedly different from understanding phrases and sentences in one's native language.

Again, surprisingly, this question is also one which has not provoked a great deal of research interest. In spite of the fact that other aspects of the performance of non-native speakers have been studied in some depth - for example the structure of the bilingual's lexicon, aphasia in bilinguals, developmental patterns in bilingualism, and so forth, - very few studies of bilingual's use of syntax have been reported. The most recent review of this field (Albert and Obler (1978)) mentions no more than a handful of studies of this type. This work will be discussed in detail in the chapters that follow.



### 1-4-3 Syntactic Structure in a Second Language

The paucity of studies that have looked at this question seems to indicate that there is a widely held consensus view that there are no major differences between native speakers and learners of a language in this regard. None of the major text-books in Applied Linguistics discusses the issue at all. Even in Leeson (1975) whose discussion of the component factors in fluent behaviour of second language speakers contains a number of sections on syntactic effects of the sort we are considering, all the data discussed is drawn from studies of native speakers, and there is no discussion of learner's performance except in terms of vague generalisations. Nowhere does Leeson explicitly consider the possibility that performance in a second language might be radically different from performance in a first language.

Where experimental work on foreign language learners has been undertaken, however, the results tend to indicate that this consensus view is false, and to support the claim that there are major differences between learners and native speakers which are specifically related to syntactic processing.

Two main theoretical positions can be distinguished in the literature. The stronger of these positions, and in some ways the more interesting claim comes from a series of studies by Macnamara (1967, 1970, Macnamara et al 1968). This work will be discussed in depth in Chapter V, and so only a bare outline will be presented here. Macnamara's claim, based on studies of word and sentence recognition, is that syntactic structure plays no part in facilitating the performance of foreign language learners. All things being equal,

sentences appear to be perceived no more quickly than scrambled lists of words in the foreign language, whereas with native speakers one would normally expect to find that sentences were perceived much more easily than scrambled word lists. Macnamara interprets this result as showing that foreign language material is processed one word at a time, without reference to the hierarchical structure of the sentence. As Albert and Obler (1978) report:

"Thus, it would appear that for English (the native language) knowledge of linear syntactic probabilities facilitated rapid reading, whereas for French (the second language) reading went on word by word." (p207).

Independent support for this claim comes from a further experiment by Trim (1970). In this study, foreign language learners were asked to memorize very long sentences (30 words) in their foreign language, and the recall patterns of this learner group were subsequently compared to those of a group of native speakers. Trim claims that markedly different recall patterns were found for the two groups of subjects, in that besides the expected quantitative differences, there were also clear qualitative differences in recall style. For the native speakers, syntactic structure and recall were closely linked, in that where words were omitted from the recall reports, they tended to be either isolated words of secondary importance - chiefly adjectives or adverbs - or entire constituents such as a whole relative clause or prepositional phrase. No such structure-related patterns emerged for the learners. Trim reports that the learner group produced marked serial position effects, the beginnings and particularly the endings of the sentences being well recalled, and the recall patterns approximating closely to the familiar serial recall curve found in studies of verbal learning. Serial position effects of this sort are common and easily

obtainable phenomena, but they are usually associated with recall or recognition of unstructured lists of items, such as sequences of unrelated nouns, or long pseudo-words made up by choosing letters at random, and it is very unusual to find serial position effects where sentences are used as the stimulus material.(e.g. Deese and Kaufman (1957)). The obvious inference from this finding is that the foreign language sentences are being treated as though they consisted merely of lists of unrelated words - that is as though their syntactic structure was a property entirely irrelevant to this reproduction task. We shall refer to this claim in future as the Word List Conjecture.

The second coherent position that emerges from the literature is less contentious than the word-list conjecture, but at the same time it is vaguer, and much more difficult to use to make clear and unambiguous experimental predictions. We shall refer to this second position as the Reduced Redundancy Conjecture. This term implies that there is a single well-thought out position here, but in fact, there are a number of separate strands which we are lumping together under this label. Each strand uses a slightly different terminology, and a slightly different conceptual framework, but there does not seem to be any fundamental incompatibility between the strands.<sup>3</sup>

The best-known of these sources is the work of Oller and his associates on Cloze Tests. Oller has argued (e.g. Oller (1973)) that the reason why learners score poorly on Cloze Tests is that these tests involve a particularly important aspect of normal verbal behaviour: the use of contextual information as an aid to making hypotheses about the text one is perceiving. Oller suggest that the chief mechanism that underlies all language behaviour - "thinking, understanding, speaking, reading and writing" - is a "grammar of expectancy."

This term is not clearly defined, but in support of his claim, Oller quotes Goodman:

"Research has demonstrated that the reader does not process print sequentially, but rather in a manner which reflects his use of language at every opportunity. Expectancies about syntax and semantics within context lead to hypotheses which can be confirmed (or disconfirmed) with only a small portion of the cues in the text."

Oller goes on to argue:

"In fact it has been shown in repeated studies that Goodman's observation applies in the case of every observable aspect of language use. The element of expectancy can be shown to affect not only the visual processing of language, but auditory processing and speaking as well....." (p112)

Oller then discusses the importance of time constraints in language processing and suggests that:

"In a test of listening comprehension, a student of a second language may have difficulty in formulating expectations and matching the incoming input to those expectations rapidly enough." (p114)

A similar type of constraint operates in reading tasks if fairly long sequences of dependent structures are involved - as is usually the case in Cloze Tests. Oller does not discuss what it is that causes the learner's "grammar of expectancy" to be defective, though he suggests that limitations on short term memory may be at least partially responsible for this shortfall. This idea is one which will reappear in Chapter 2.

The second source for the reduced redundancy conjecture is the work of Spolsky and his associates on noise tests (Spolsky et al (1968)). Spolsky's work is much less well-known than that of Oller, and considerably less influential, though this seems to be due principally to the technical problems of working with white noise when compared with the simplicity of the Cloze Procedure. Spolsky

explicitly explains the poor performance of learners on his tests in terms of reduced redundancy in foreign language material, arguing that while native speakers can cope with degraded messages because they can recover obliterated information from other parts of the message, learners appear to be unable to do this, and need a much clearer image if they are to achieve understanding:

"Messages in normal language can be understood even though a good proportion of them is omitted or masked; or, in other words, every message contains many elements..... that can be omitted without leading to a breakdown in communication. But if we give these distorted messages to someone who doesn't know the language well, we find that there is a considerable difference. He needs the full normal redundancy, and at times even that is not enough. Note how when we speak to someone whose native language is different, we tend to speak more slowly, more clearly, with added gesture and frequent repetitions." (p386).

Like Oller, Spolsky offers no clear evidence that might explain why learners cannot cope with messages with added noise, nor why it appears necessary for learners to be supplied with messages of exceptional clarity in order to avoid breakdown of communication. Both these sources are informal, in the sense that they do not formally test their claims in any way. Rather, they use the idea of reduced redundancy as an explanatory notion, with which to account for the poor performance of the learners studied. Some more formal support for the conjecture comes from the work of Triesman, whose experiments are discussed more fully in Chapter 2. Triesman's approach to redundancy is a more technical one than is found in the work discussed so far. It makes use of measures of entropy, and it is argued that the entropy value of foreign language material is consistently higher than the entropy value of the same material

presented to groups of native speaker subjects. Triesman's work is the only investigation which amounts to a formal test of the reduced redundancy conjecture, but as we shall see in Chapter 2, there are serious problems with her entropy measurements, and this makes it very difficult to take the results of this work at face value.

The discussion that follows will be principally concerned with testing the word list conjecture. This is mainly a matter of experimental convenience, rather than the result of serious theoretical considerations. As we have seen, the evidence supporting neither of these conjectures could be considered really compelling, as the amount of research work that has been carried out in this area is far too limited in scope for anything like a coherent theoretical picture to have emerged.

There are, however, a number of difficult practical problems with the reduced redundancy conjecture that make it difficult to work with as an experimental hypothesis. The principal difficulty is that the conjecture is not so much a single coherent explanation of why learners perform badly on foreign language material, but rather an overall description of their behaviour, which encompasses a very wide range of possible explanatory principles. Indeed, the Word List conjecture could really be considered as a special case of the Reduced Redundancy Conjecture, in which the notion is taken to an extreme position. The Reduced Redundancy Conjecture is based on a claim made on general psychological grounds that redundant material is handled more efficiently than less redundant material, and it is shown that learners handle foreign language material less effectively than material presented in a first language. It is then argued that the material must therefore be effectively less redundant for

the learner than for native speakers. The structure of this syllogism is faulty, of course, as the conclusion does not logically follow from the premises. In principle there could be many reasons why foreign language texts are inefficiently processed which have nothing to do with redundancies of any kind. However, even if we allow that the claim might be true, there are still a large number of particular factors which could account for an apparent decrease in redundancy. These factors could interact in complex ways, and where two different types of experiment produce results that appear to support the reduced redundancy conjecture, there is no guarantee at all that these results are ascribable to identical causes. If we consider again the work of Spolsky and of Oller, for instance, it seems plausible to suggest that a major factor contributing to the poor results of learners on Spolsky's noise tests is a general difficulty in identifying words in degraded conditions. It would be difficult to argue convincingly, however, that identification of words is a major problem for learners in Oller's Cloze task where the stimulus material is presented in written form. This lack of precision in the Reduced Redundancy Conjecture makes it very difficult to set up experiments which would allow us to find conclusive evidence against it, so that the conjecture appears in some respects to be relatively invulnerable. The only exception to this is a small number of experimental paradigms which make use of a global measure of redundancy, treating this variable rather like other more objective properties of sentences such as sentence length. In experimental settings of this type it is possible to make a general prediction that learners will be less effective performers than native speakers, and in theory this prediction could be falsified by finding no differences between native speaker and learner

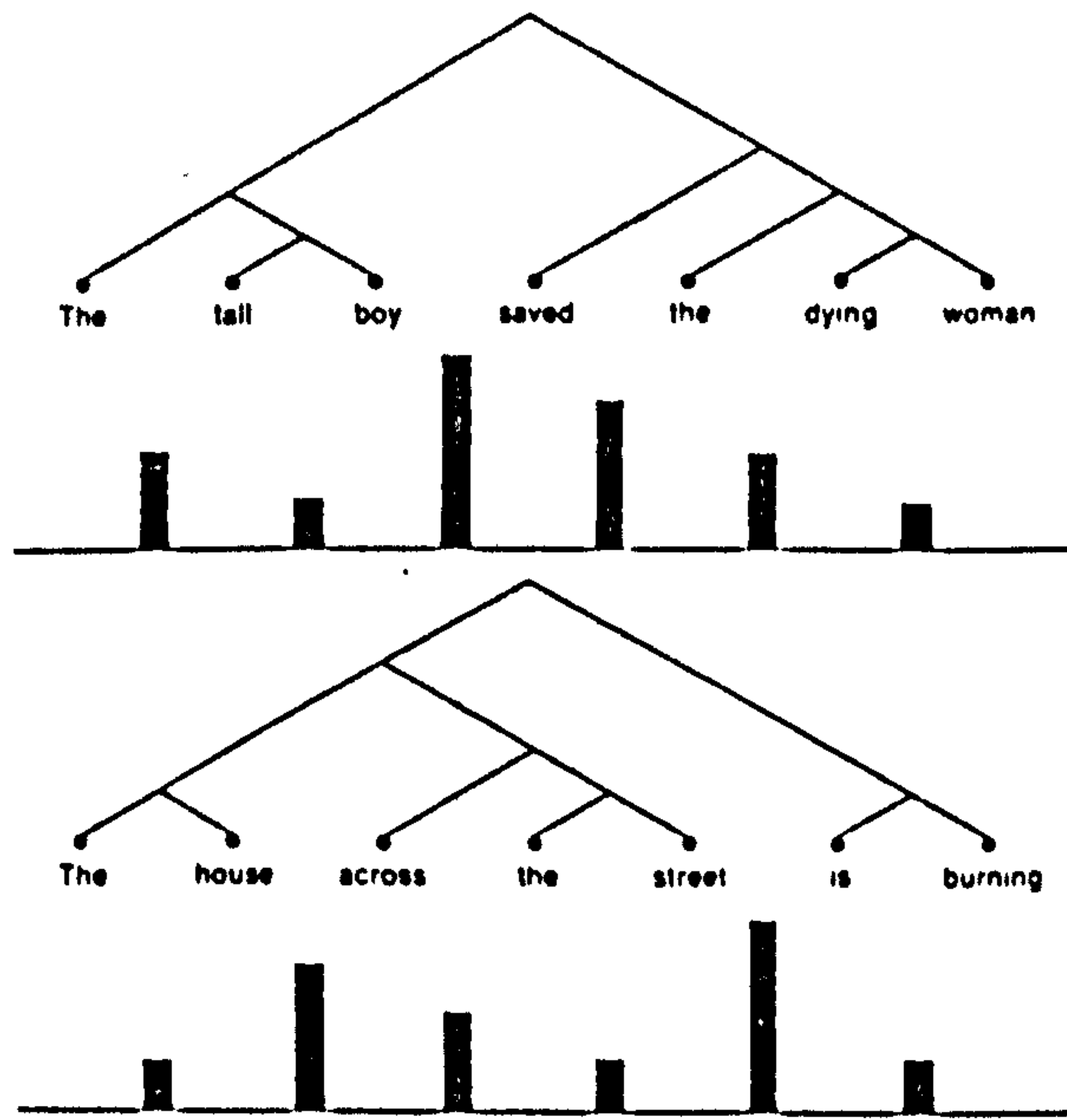
subjects. In practice, however, one does not normally set up experiments which look for evidence of no difference between groups, as such claims are equivalent to null hypotheses. In any case, given that there is a well-established general performance deficit on most foreign language tasks (e.g. Long and Harding-Esch (1977)) evidence which supports the Reduced Redundancy Conjecture in its crude form is fairly easy to come by, and not particularly illuminating.

If we move away from general measures of redundancy, and consider more specific experimental tasks, it becomes much more difficult to decide what predictions the Reduced Redundancy Conjecture would lead us to make. Consider, for example, Johnson's Transitional Error Probability experiments (Johnson (1965)) referred to briefly above. In these experiments, subjects are required to learn lists of sentences, and at various stages in their recall attempts, note is taken of errors made in recall. Johnson noted that for native speaker subjects, the likelihood of making an error depends to some extent on the syntactic structure of the particular sentence being recalled, in that errors are much more likely to occur at the boundaries of constituents than within them, and that more errors occur at major constituent boundaries than at minor ones. This pattern of results is shown in Figure 1-1 on the following page.

Now, imagine that we are subjecting learners to this task, in order to test the reduced redundancy hypothesis. What predictions does the conjecture lead us to make? The principle prediction is that learners would make more errors, but within the limitations imposed by this constraint, any error pattern would support the conjecture, whether it showed any similarity to the error curves of the native



Figure 1-1. Transitional Error Probabilities for two sentences  
 After Johnson (1963)



speaker or not. Clearly evidence that supports the claim covers such a wide range of possible events that it is not particularly informative when it is found.

In contrast to this, the word list conjecture does make very clear, explicit and unambiguous predictions about the performance of native speakers. For instance, in the case of Johnson's paradigm, the word list conjecture would lead one to hypothesize that error patterns made by non-native speakers would be more likely to be influenced by serial position than by any syntactic structure, and that this should cause non-native speakers to produce similar transitional error patterns for a variety of different sentence types which could be shown to produce different error patterns in native speakers. Clear predictions of this type are an obvious tactical advantage when undertaking experimental work, and it was therefore deemed wiser to probe the word list conjecture, rather than to work with

the inherently vaguer claims based on the Reduced Redundancy Conjecture. It will become apparent from the discussion that follows in the subsequent chapters, that the word list conjecture leads to some false predictions - although some evidence which supports a weaker version of the conjecture will also be presented. This data will enable us to derive a number of more specific hypotheses about the way non-native speakers behave in a second language.

CHAPTER 2.        EXPERIMENTS WITH STATISTICAL APPROXIMATIONS.

2-1    Introduction

2-2    General Background

2-3    Statistical Approximations and Contextual Restraint

2-4    Statistical Approximations and Foreign Language Learners

2-5    Experiment 1

2-6    Experiment 2

2-7    Discussion and Comments

2-8    Conclusions

## 2-1. INTRODUCTION

The principal object of this chapter is to establish whether structural constraints of any kind can be shown to have an effect on the performance of non-native speakers of a language. For this reason, we have chosen to study an experimental task in which "structure" is rather loosely defined, and treated as a global property of the strings of words that are used as stimuli. The fine detail of this structuring will not concern us here, although some experiments in which this fine detail of structure is important will be considered in subsequent chapters.

The task studied in this chapter is an immediate recall task, and the stimulus material used consists of a set of statistical approximations to French - i.e. strings of French words which show varying degrees of structurability. This combination of task and materials was chosen because there is a long tradition of using materials of this type in order to answer questions about the ability of groups of subjects to handle syntactic structure, and though the immediate recall task is a fairly insensitive one, in the sense that it can show us only gross differences in the performance of experimental groups, as we shall see in the later sections of this chapter, it is sufficiently sensitive to show that there are some substantial differences between native and non-native speaker subjects in their ability to handle materials of this sort.

The chapter falls in eight sections. Section 2-2 discusses some relevant background work on the performance of non-native speakers of a language on memory tasks; 2-3 describes statistical

approximations, and the experiments in which they have been used. Section 2-4 discusses the work of Triesmen - the only previous use of statistical approximations with non-native speaking subjects. Sections 2-5, 2-6 and 2-7 present an account of some new experimental evidence reported here for the first time, and the implications of this evidence are drawn out in Section 2-8.

## 2-2. GENERAL BACKGROUND.

### 2-2-1 Short term memory and sentence processing

Some justification for using an immediate recall task as the starting place for our investigation of the performance of foreign language learners is perhaps necessary here. Immediate recall of strings of words is not the most natural of language tasks, nor is it an activity that normal speakers engage in with any frequency. Furthermore, it might be argued that immediate recall tasks bear little relationship to normal language use, and to single out this one aspect of linguistic behaviour as the object of a laboratory study necessarily introduces distortions which will make it difficult to generalize the results to more natural, less artificial tasks.

Nevertheless, there are a number of reasons why immediate recall tasks seemed to be a good place to begin this study. Firstly, memory span as measured by immediate recall of word lists is one of the very few performance characteristics of foreign language learners that has received any attention whatsoever, and it seemed wise to begin these investigations in an area that was not wholly uncharted. This previous work is discussed more fully in section 2-2-2. Secondly, there is an enormous body of experimental data concerning immediate recall tasks for both normal native speakers of English, and a wide range of non-normal speakers, and the basic phenomena in this field are among the best established results of experimental psycholinguistics. This means that it is fairly easy, in principle at least, to ascertain when an experimental group of subjects is behaving in an abnormal fashion, and it becomes

relatively easy to generate plausible explanations for such abnormal behaviour when it is found. The third reason for choosing immediate recall as a starting point is that though the task may appear to be widely removed from normal language behaviour, it is actually less far removed than many other laboratory tasks, and involves processes that play a central part in normal language handling.

All modern theorists of word recognition and of sentence processing appear to be agreed that some form of short-term storage is an essential component of any language processing model, although there is some dispute about the parameters of such a component, or whether there is only a single component of this type, or a number of similar components with slightly different characteristics each acting at various points in the comprehension process. The chief function of this component or components is to act as a buffer store, holding incoming information long enough for it to be decoded, and for each component part of the message to be related to other relevant parts. The principal characteristics of stores of this type are the rapid fading of their contents, and their capacity to hold only a limited amount of information.

Although there is considerable disagreement over the actual capacity of the short term storage system(s), and over how this characteristic can be accurately assessed, it is generally agreed that this storage capacity is not large. Miller (1966) quotes a figure of 7 plus or minus two items, but even this low figure has been challenged more recently, and Simon (1974) suggests that Miller's estimate is almost certainly an overestimation of the true capacity.

It is widely recognized, however, that whatever the limitations on short term storage are, they can be easily circumvented, particularly if the material to be remembered can be restructured in some

more efficient form. Smith (1954) for example, showed that very long strings of binary digits - up to forty items - could be remembered easily if the strings were segmented into blocks of 4 or 5 digits and these blocks were recoded by being given a single syllable name. Bousfield (1953) showed that apparent memory span could likewise be increased if semantic or phonological criteria could be used to group otherwise isolated items in a word-list.

Explanations of how such grouping takes place, and of the exact mechanisms by which the apparent limitations on memory span can be overridden vary widely, depending on the type of model under consideration. For the purposes of discussion here, we will adopt Miller's (1956) chunking model, in preference to other more complex explanations. Miller's account is not a wholly adequate account of the phenomena, but it is an elegant and simple metaphor, and provides a useful framework which will allow us to make some preliminary observations about the performance of non-native speakers.

Miller proposed that short term storage capacity could be viewed in spatial terms as a finite number of slots - usually about seven - with a strictly limited capacity. Each of these slots could be filled by one item of information from the material to be remembered. In the case of a list of unrelated words, each slot would be filled by a single word, and if the number of words to be remembered exceeded the number of slots available, then some of the material would fail to be recalled. In some circumstances, however, it might be possible for items to be grouped together into chunks of more than one word, each chunk occupying only a single slot. Miller argues that this idea accounts for Smith's ability to remember long sequences of binary digits; the recoding of blocks of digits into single code words means that a large number of items are chunked into



a single more efficiently coded item, which can fit comfortably into one of the available slots. Storage capacity in terms of slots remains unchanged, of course, but apparent storage capacity, measured in terms of items recalled, increases. Miller suggests that Smith's overt recoding strategy is akin to a covert recoding process that takes place automatically whenever we have to remember material for which a coding system is available.

Normal language processing appears to involve a covert recoding system of this sort. As we have seen earlier, few native speakers are limited in any obvious way by their absolute memory span when it comes to handling ordinary speech or written language. Most native speakers can reproduce verbatim a sentence of twenty or thirty words a figure that greatly exceeds memory span for isolated words.

Furthermore, rapid speech and rapidly presented sentences in the visual modality are both easily handled, even when word lists presented at equivalent speeds would be impossibly difficult.

(Garvey (1953), Foulke (1971), Forster (1973)). It seems therefore, that native speakers must have at their disposal some form of processing strategies that enables them to handle incoming messages in terms of units much larger than the individual words that make them up, and thus to reduce the perceptual load of the messages. Miller suggests that syntactic structure might form the basis of such a recoding strategy, with incoming messages being chunked in terms of their constituents.

## 2-2-2. Memory Span and Foreign Language Learners.

It has been known for some time that memory span is an important component in foreign language ability. Both the main Language Aptitude Tests, Pimsleur (1966) and Carroll and Sapon (1959) contain tests of memory span which receive a considerable weighting in the overall aptitude score. In spite of this, however, there is not a great deal of evidence available about the memory span of foreign language learners. A number of studies (e.g. Cook (1978)) have indicated that memory span for foreign language material is rather smaller than it is for material in the native language, though even this relatively clear finding is in doubt. Lambert, Ignatow and Krauthamer (1968), for example, report that advanced learners studied by them had a larger span in the foreign language (French) than in their native language (English).

An important source in this area is Lado (1965) where a number of studies concerning memory span in second languages are reported and discussed. Lado reports two findings which are of particular importance for this study. Firstly, he reports a simple digit span experiment in which native speakers of English were tested on lists of English and Spanish digits. The results showed that subjects' memory span in English was reliably higher than in Spanish, the mean difference amounting to 1.6 digits. Secondly, Lado reports a study by Glicksberg, again a simple memory span experiment using English and Spanish subjects. Lado reports that a reliable foreign language memory span deficit was found when the material to be remembered consisted of isolated lists of words. He further reports that this deficit increases in size when the stimuli consist of "connected

material", i.e., presumably, sentences. Details about the size of the increase are not provided, but it seems that we are dealing with a straight-forward interaction here: memory span in both languages increases with structured material, but the increase is comparatively larger in the native language case.

Both these studies have been widely reported, and their results have been usually accepted at face value, (e.g. cf. Cook (1978) or Bennett (1974) p199). It seems to me, however, that both sets of data need to be treated with some caution. It is not clear, for instance, how far digit span tests in a foreign language are reliable indicators of how other types of material might be handled. Digits in one's native language are highly overlearned and relatively frequent items. This, tends to increase the ease with which they are handled. It is well known, however, that digits cause some considerable difficulty even to highly advanced learners of a foreign language, who often stumble over large numbers, for example, long after most other major difficulties have been overcome. It seems unwise therefore, to accept this finding without further corroborating evidence. A further point is that Spanish digits are typically disyllables, while English digits are typically monosyllables, and this too would lead one to expect slightly worse results in Spanish. (cf. Dornic (1969)). Whether these two factors would account for the rather small difference of 1.6 digits, or whether there is a genuine memory span deficit here, is thus not clear.

The word length consideration is a factor which also needs to be taken into account in Glicksberg's data, since Spanish words are typically longer than English ones but there is a further problem here too. Glicksberg compares lists of words and sentences, but this is not really a fair comparison, as sentences have a number of characteristics other than the presence of syntactic structuring that can

affect scores on a memory span test. Sentences typically contain a high proportion of function words - articles, prepositions, etc. - which are usually short, and are more frequent than full lexical words such as nouns and verbs. This means that one would expect sentences to be better recalled than lists of words anyway, and so it is not clear that the improvement reported for the foreign language sentences in this experiment represents a real structural effect, after all. A much better test of the syntactic effect is to use either jumbled sentences, where the order of the constituent words is changed, but other characteristics remain the same, or alternatively to use lexically dense sentences in which length and frequency factors are reduced to an absolute minimum by the exclusion of function words. Neither of these techniques is wholly satisfactory, however.<sup>4</sup> It seems then, that though at first sight Glicksberg's data suggests that learners do respond to syntactic structure in foreign language sentences, this finding is sufficiently doubtful that it does not constitute a serious refutation of the word list conjecture that learners do not respond to syntax, and that further work in this area is warranted.

## 2-3. STATISTICAL APPROXIMATIONS AND CONTEXTUAL CONSTRAINT.

### 2-3-1 Statistical Approximations.

One of the standard experimental techniques for investigating the effects of linguistic structuring on performance is the use of statistical approximations. These are strings of words in which the structural redundancies which characterize normal language are systematically manipulated in a manner which is described in detail below. It is this type of material that is used in the experiments reported in this chapter.

## 2-3-2 Construction of Statistical Approximations.

Statistical approximations were first used by Miller and Selfridge (1950) though a similar idea based on letter frequencies had been suggested already by Shannon (1948). Miller and Selfridge distinguish between different orders of approximation on the following lines:

- a 0-order approximation is one which takes no account of the sequential dependencies that normally characterize English, i.e., it is basically a list of words based on a random sampling from a standard word list such as a dictionary.
- a first order approximation is similarly a list of words, but a list which reflects the frequency of occurrence of the items in sentences of the language - i.e. a list of words based on random sampling of continuous text.
- higher orders of approximation are word lists that reflect the probability of occurrence of longer sequences of words, for example, of word pairs or word triplets. Theoretically, these higher orders of approximation could also be constructed by appropriate sampling techniques, but this becomes very difficult and tedious for longer sequences, and in practice higher order strings are constructed as follows:

A Subject, S1, is provided with N words and asked to make up a sentence in which these N words appear adjacent to each other. Let us assume here that N is 2, and let us designate the two words A and B. In the sentence that S1 produces, these two words will be followed by a third word which we will designate C. A second Subject, S2, is now asked to use B and C consecutively in a sentence.

The word that follows C in his sentence, let us call it D, is now used to make a context for a third subject, S3, who is required to use C and D consecutively in a sentence. D and E (the word following D in S3's sentence) are next presented to S4 to be included into a further sentence, and this process of removing the first word of the stimulus string and adding at the end of the string the word that followed it in the sentence is repeated indefinitely. Eventually these isolated responses are collected together into a single string, ABCDEFGH..... Such a string of words, with two words of context given to each subject who is asked to provide a third word is called a third order approximation (to English). A string formed in the same way from stimulus strings of three words context and one additional word provided by the subjects is called a fourth order approximation, and, in general, an Nth order of approximation is generated from N-1 words of context. A sample derivation of a third order approximation which illustrates how this procedure works in practice is shown in Figure 2-1 on the following page.

For a further discussion of the method see Miller and Selfridge (1950), and for some critical comments see Meara (1975).

Fig.2-1 Derivation of a third order approximation to English.

- a) the man  
the man saw the dog.
- b) man saw  
the man saw his wife coming.
- c) saw his  
the boy saw his kite fall to the ground.
- d) his kite  
his kite was shaped like a dragon.
- e) kite was  
the kite was red and yellow.
- f) was red  
the sky was red and lowering.
- g) red and  
roses are red and scented.
- h) and scented  
her handkerchief was made of lace and scented, while  
her dress was black and severe.
- i) scented while  
some flowers are scented while others are not.
- j) while others  
it is wrong that some people live in luxury while  
others are hungry.
- k) others are  
the others are coming.
- l) are coming  
if you are coming, bring a bottle.
- m) coming bring  
since you are coming bring her with you.
- n) bring her  
he said he would bring her a single red rose.

The completed passage would now read:

the man saw his kite was red and scented while others  
are coming bring her a .....



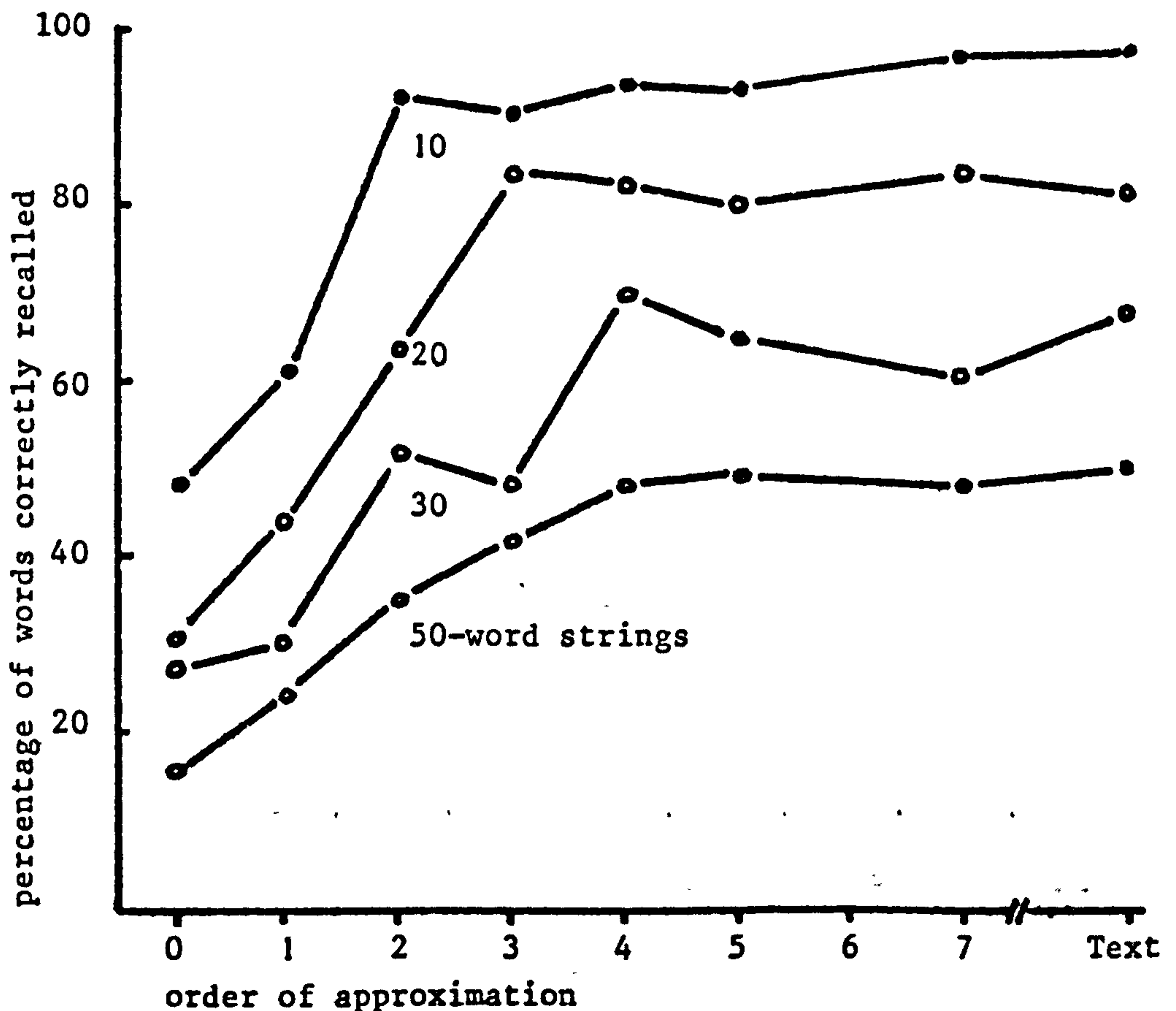
2-3-3 Miller and Selfridge's Experiment and the standard results for native Speakers.

> Miller and Selfridge constructed a set of stimuli consisting of word strings of different lengths at various orders of approximation to English. Their subjects heard each string and were required to recall what they had heard at the end of each string.

-> Miller's and Selfridge's results are shown below in Figure 2-2. No statistical handling of these data is offered, but the results suggest that Subjects handled the higher orders of approximation with rather more ease than the lower orders. It also appears that there is little difference between the third and higher orders of approximation, where the recall scores seem to reach an asymptote.

Figure 2-2.

Percentage of words of the lists of different lengths that were correctly recalled at the various orders of approximation to the statistical structure of English. (After Miller and Selfridge (1950)).



## 2-3-4 Statistical Approximations: interpretation and further work.

Miller and Selfridge explained their findings in terms of inter-item associations. They argue that passages are easy to recall if they preserve the short-range associations that are familiar to the subjects. Long-range associations between items appear to be of importance only with long word strings.

Subsequent research tended to confirm Miller and Selfridge's original findings. cf., for example, Sharp (1958), Postman and Adams (1962), Richardson and Voss (1960), Tulving and Patkau (1962) etc.... There was, however, some discussion over the nature of the asymptote found with higher order approximations, and a variety of scoring techniques were devised to show monotonic relationships between order of approximation and recall. cf., for example, Marks and Jack (1952), and Johnson (1968).

The findings were also subjected to a number of reinterpretations. These fall into two main groups. The first group argued that statistical approximations differ from each other in objective ways, and that the improved performance of subjects on higher orders of approximation can be attributed to these differences. Typical of this group of commentators is Deese (1961) who argued that the difference found between 0-order and 1st-order approximations can be largely accounted for in terms of the guessing behaviour of subjects. Also typical of this approach are Salzinger, Portnoy and Feldman (1962) and Triesman (1965). The former discuss the degree of redundancy in statistical approximations which they estimated on the basis of Cloze-type tests. Their results showed that predictability increases with order of approximation, though the increases from one order to the next are not uniform. Triesman's approach is very similar to this.

She calculates the degree of entropy of the original passages used by Miller and Selfridge, and shows that this value decreases with order of approximation (though in a later paper Harison, Triesman and Murray (n.d.), it is argued that this last claim is something of an oversimplification).

The second group of comments suggested that the differences found in the level of performance are due to the subjects being able to use different strategies for handling the material. The strategies used on higher orders of approximation are more efficient than those adopted for lower orders, and so allow more material to be stored and recalled. Typical of this group of comments are Deese and Kaufmann (1957), who showed that subjects changed their recall patterns with higher orders, in that the normal serial recall curve associated with free recall of unstructured material was replaced by a striking primacy effect, as the subjects tended to recall the items in the higher order strings in the same order as they heard them. This paper is discussed in more detail below.

These two approaches are not, of course, incompatible. They are synthesized in Johnson (1968) who suggests that it is unlikely that the statistical approximation effect is the result of a single, undifferentiated factor. More probable is that several factors are at work, and that these operate differentially at different orders of approximation. He suggests that the greater predictability of high orders of approximation is what allows subjects to adopt their more efficient strategies.

## 2-4. STATISTICAL APPROXIMATIONS AND FOREIGN LANGUAGE LEARNERS.

### 2-4-1 General Considerations.

In the previous section we discussed the standard findings for native speakers of English on recall of statistical approximations. The question which now arises is what predictions are made about the performance of non-native speakers on this task by the word-list hypothesis?

The principal prediction is that learners' performance ought to be totally unaffected by the characteristics of higher order approximations, and a result along these lines would provide strong support for the claim that learners operate simply in terms of individual words when they handle strings in their foreign language, and that the syntactic structure of these strings does not play a major part in the way they are processed by non-native speakers. This prediction is shown in diagram form in Figure 2-3a. The figure actually shows two possible patterns of learner performance. The lower line shows no difference between the various orders of approximation as outlined above. In practice, however, one might expect that a 0-order approximation would be harder to handle than any higher order approximation. This is because higher orders take account of the frequency of occurrence of individual words, and frequent words are generally shorter than infrequent ones (cf. Zipf (1949)). One of the chief characteristics that differentiates 1st-order and 0-order approximations is that the former contain a high proportion of short function words - articles, prepositions, pronouns, etc. which should have the effect of making the 1st-order approximations considerably easier to handle than their 0-order counterparts. This outcome is

Figure 2-3.

for explanation see text.

native speaker scores are hypothetical  
and % recall level is arbitrary.

■ native speakers  
○ learners

Figure 3a

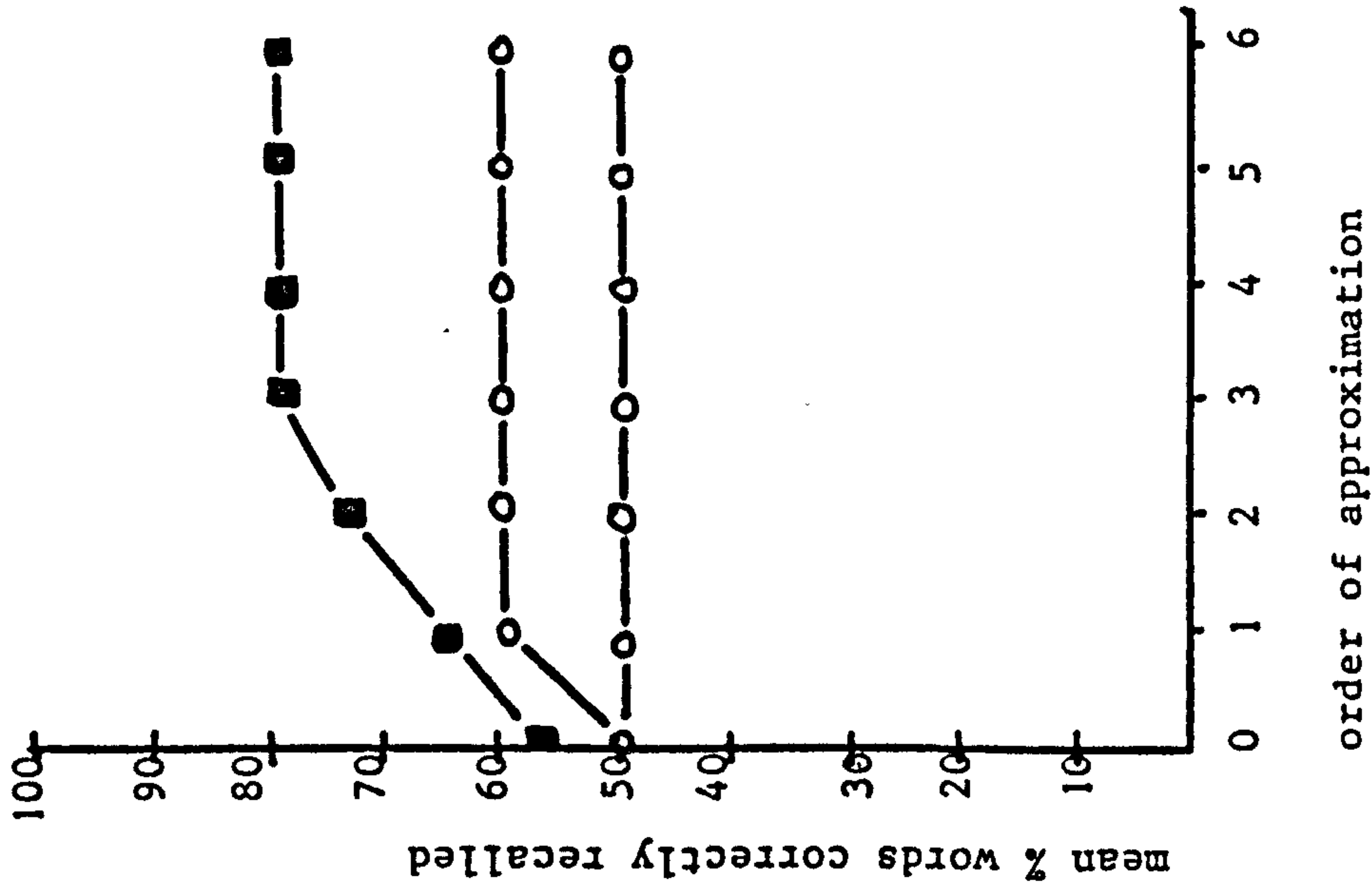


Figure 3b

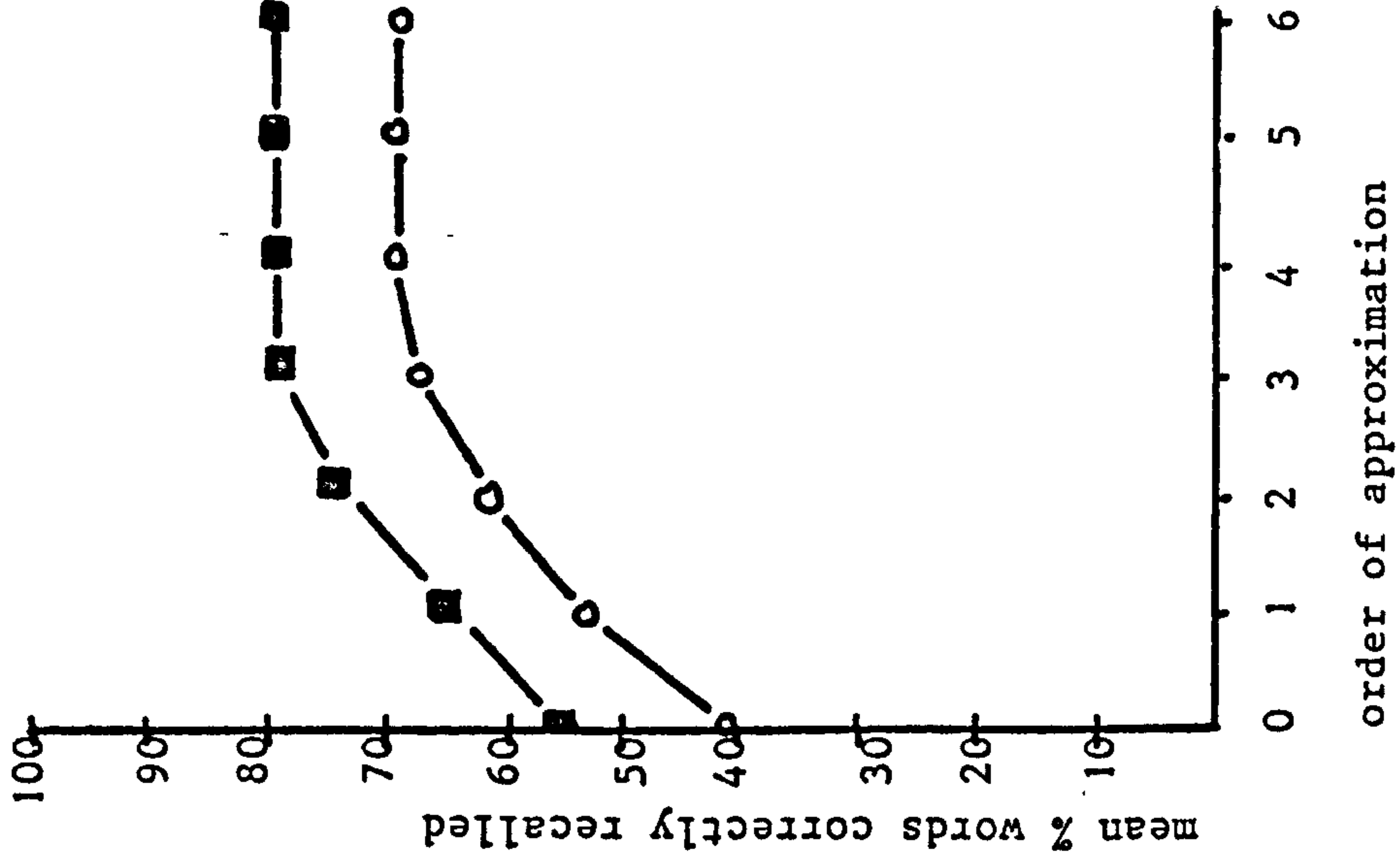
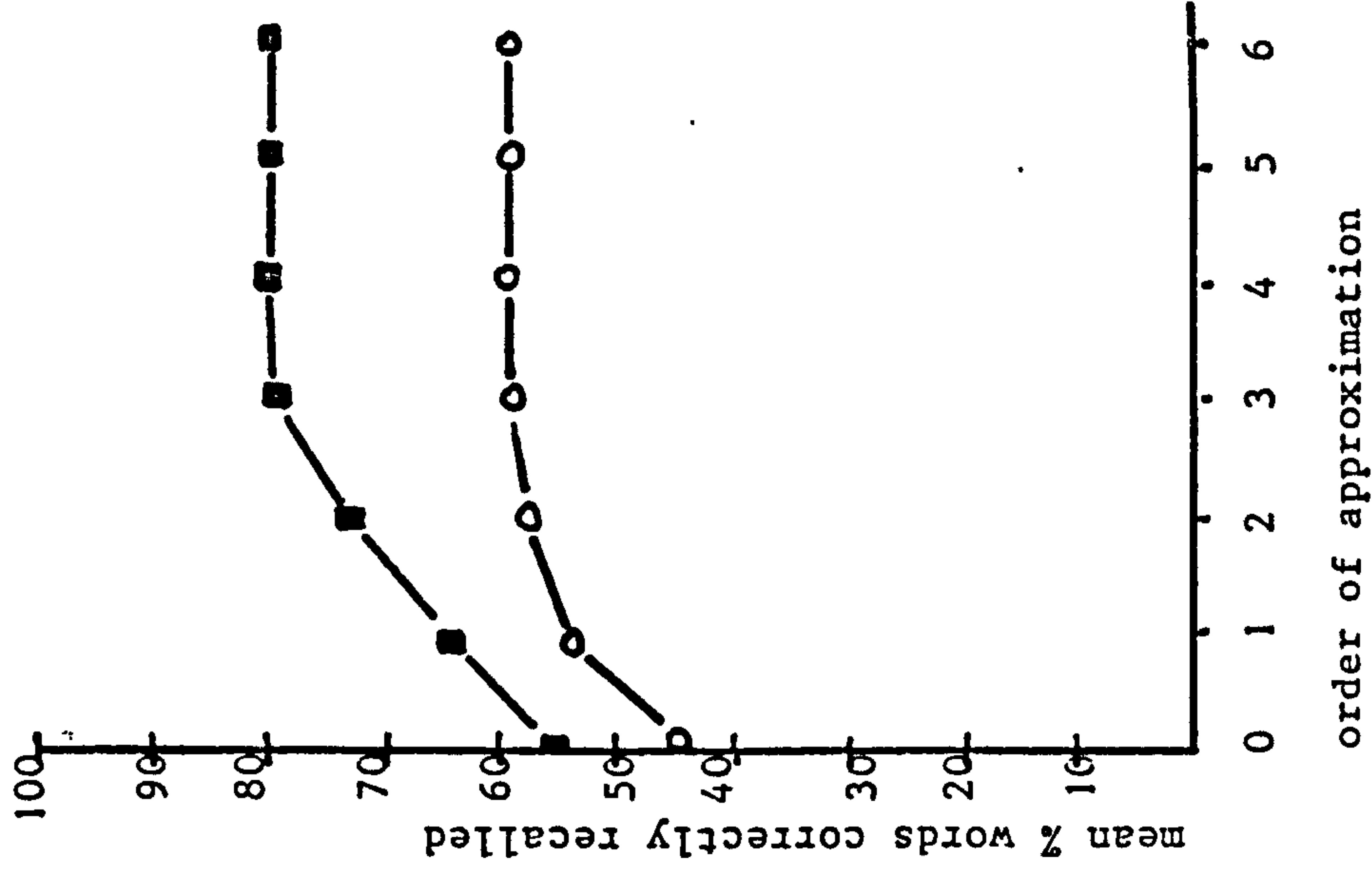


Figure 3c



also shown in Figure 2-3a, and is represented by the upper line of circles.

Two other hypothetical outcomes of an experiment with statistical approximations are shown in Figure 2-3 for the purposes of illustration. Figure 2-3b shows the non-native speaker scores running broadly parallel to the scores of the native speakers, the difference in performance between the two groups remaining fairly constant in absolute terms. A result of this sort would clearly count as strong evidence against the word list conjecture since it implies that the difference between native speakers and learners are not affected in any way by varying the syntactic cohesion of the input strings. Rather, there is a constant difference in level between the groups, which would have to be accounted for in terms of some non-structural factor such as the familiarity of the stimuli.

Figure 2-3c represents a whole set of possible outcomes all of which involve interactions of a fairly complex sort. In Figure 2-3a, the non-native speakers are envisaged as improving their scores on higher order approximations, but to a far smaller extent than is normally the case with native speakers, so that the two sets of scores gradually diverge. An outcome of this sort would also count as evidence against the word list conjecture, since the learner's performance is clearly affected by order of approximation. At the same time, however, such a result would indicate that only a very small improvement in recall is available to learners, and this would point to some serious limitation on non-native speakers' ability to use structural information in a foreign language.

## 2-4-2 Previous work with learners.

The only work that I have been able to trace that looks at the performance of non-native speakers on statistical approximations is that of Triesman (1965). Triesman's results are not directly comparable with the results to be reported in the next section, however, since she uses not an immediate recall task but the more complex task of shadowing. Nevertheless, Triesman's results are important since they suggest that learners' performance is affected by syntactic structure in much the same way as the native speakers' performance. Her results seem therefore to disprove the predictions made by the word-list conjecture.

Triesman asked native speakers of French and English to shadow and translate approximations to these two languages. We are not concerned with the translation task here, as it involves skills other than those involved in the perception and reproduction of strings in a single language. On the shadowing task, Triesman found that the percentage of words correctly shadowed varied with order of approximation, and more particularly, that this figure varied with the rate of presentation of information measured in bits per second. Since low orders of approximation are less redundant than higher orders, the mean information per word is greater for low orders than it is for high orders. If the strings are presented to the subject at the same number of words per minute, lower orders will thus show a higher rate of information presentation.

Triesman argues that her results show that shadowing in a foreign language increases the effective information rate, since, while the regressions of correct responses on information rate are parallel, the scores for performance in the foreign language are consistently

worse than the scores for performance in the subjects' native language. The only exception to this is the case of bilingual subjects, where the regression co-efficients are identical. These results are clearly evidence against the word-list conjecture, which predicts that for non-native speakers syntactic redundancies should not contribute towards reducing the effective information rate. Triesman's results suggest that there is a constant information increment due to operating in one's foreign language, and that this increment is largely unaffected by syntactic factors and increased contextual constraint. The word list conjecture would lead one to predict that Triesman's regression lines ought to diverge rather than running parallel, since if learners are insensitive to syntax, the effective information rate of the higher order strings ought to be relatively greater for the non-native speakers than it is for the native speakers, since with these strings information rate is reduced because of syntactic constraints.

Triesman's results need to be treated with some caution, however. In the first place, she appears to be working with very advanced learners. Her native French speakers all seem to have lived in Britain for some time. In addition, the scores of the native English speakers on the French approximations are very similar to those of the native French group, and it seems reasonable to infer from this that her undergraduates were very fluent in French. Triesman does not discuss periods spent abroad in French speaking countries, but it is well-known that extended residence abroad can result in a high level of fluency, so it is not entirely surprising that the performance of Triesman's subjects in their foreign language should broadly resemble their performance in their native language. It is not at all clear, however, how less advanced learners might be expected to



perform on a task of this sort.

A second problem is that Triesman reports her results in terms of information rate rather than in the form that is normally used for work with statistical approximations. This gives the impression of rigorous accuracy, but the way in which the information rate figures are calculated is suspect. Triesman's figures are worked out on the basis of a Cloze procedure, a hundred subjects being asked to fill in the gaps of typescripts of the passages from which every tenth word was deleted. The resulting entropy measure is in itself a fairly crude measure, since it is based on an average of the deletions. This may be a fairly good estimate for the passages at low orders of approximation, but in high orders of approximation, a mean entropy measure makes much less sense. In higher orders there is considerable variation to how constrained any individual word can be. Some words are almost uniquely specified, while others can be almost wholly unspecified, and this means that the degree of constraint is much less uniform here than for low orders of approximation. It is not clear that sampling every tenth word does anything to resolve these difficulties. If anything, the procedure seems to make the entropy measure even less reliable. This doubt about the reliability of the entropy measures is strengthened by the fact that these measures are based on a reading test, the Cloze scores, whereas the shadowing task from which the experimental data comes uses auditory input in what is basically an aural comprehension test. It is doubtful whether the entropy measures based on the Cloze test scores accurately reflect the difficulty of guessing the next word in an auditory sequence, where only the preceding context is available, and where it is impossible to backtrack over the material already heard.

The fact that the passages for shadowing were read introduces a further set of problems that Triesman does not discuss. Written sequences can often be ambiguous, but reading such a sequence aloud will usually force one to choose between alternative readings, and so to disambiguate them. The sequence they can fish, for example, is ambiguous between they put fish into cans, and they are able to fish, but it is impossible to read the sentence aloud without making clear which meaning is intended. This means that reading the passages aloud should have the effect of reducing the uncertainty found in the written version.

A further problem arises in connection with the entropy figures for the French sequences. Triesman does not measure this at all, but assumes that the information content of the French passages is the same as that for the corresponding English passages. It is not clear how far this assumption is justified. It has been argued elsewhere (Meara (1975)) that English differs from some other languages, including French, in that it is rather less constrained at low orders of approximation. French, for example, has two classes of nouns, masculine and feminine nouns, which occur in different contexts; French also has a set of concord rules which operate over short stretches of text but have little influence over longer ones. These rules ought to lower the effective information contained in second and third order sequences where short-term sequential constraints are very important. It also ought to make the information carried by each word in higher orders more evenly distributed.

It should also be pointed out, of course, that entropy estimates based on judgements by native speakers may not necessarily correspond to the effective entropy rates for a group of learners. Entropy is not an objective characteristic of a text; it is defined in terms of

the performance of a given population. The hundred native speakers used in assessing the entropy values in Triesman's study may be assumed to be representative of the native speaking population, but since learners belong to a different population, entropy values really need to be calculated separately for them.

Figure 2-4 shows Triesman's results recalculated and presented in the more normal fashion. A detailed analysis of these results is not possible, of course, but the overall picture is confusing. Figure 4a closely resembles figure 3b, the scores for the learner group running parallel to those of the native speakers, but at a lower level, while figure 4b is closer to the outcome discussed in figure 3c - no difference between the groups on the first order list, both groups improving their performance on the more organized lists, but the native speakers showing a larger increment than the learner group. Both sets of results seem to offer evidence that contradicts the predictions made by the word-list conjecture. They are contradictory, however, in that the results on the English language passages seem to show a steady improvement on higher orders of approximation, with a slowly reducing gap between native speakers and learners, while in the French passages, the native and non-native groups are indistinguishable on the first order passage, but soon diverge and remain a constant distance apart.

Figure 2-4.

Mean % words correctly shadowed in English and French at various orders of approximation.

Figure 2-4a

English language passages

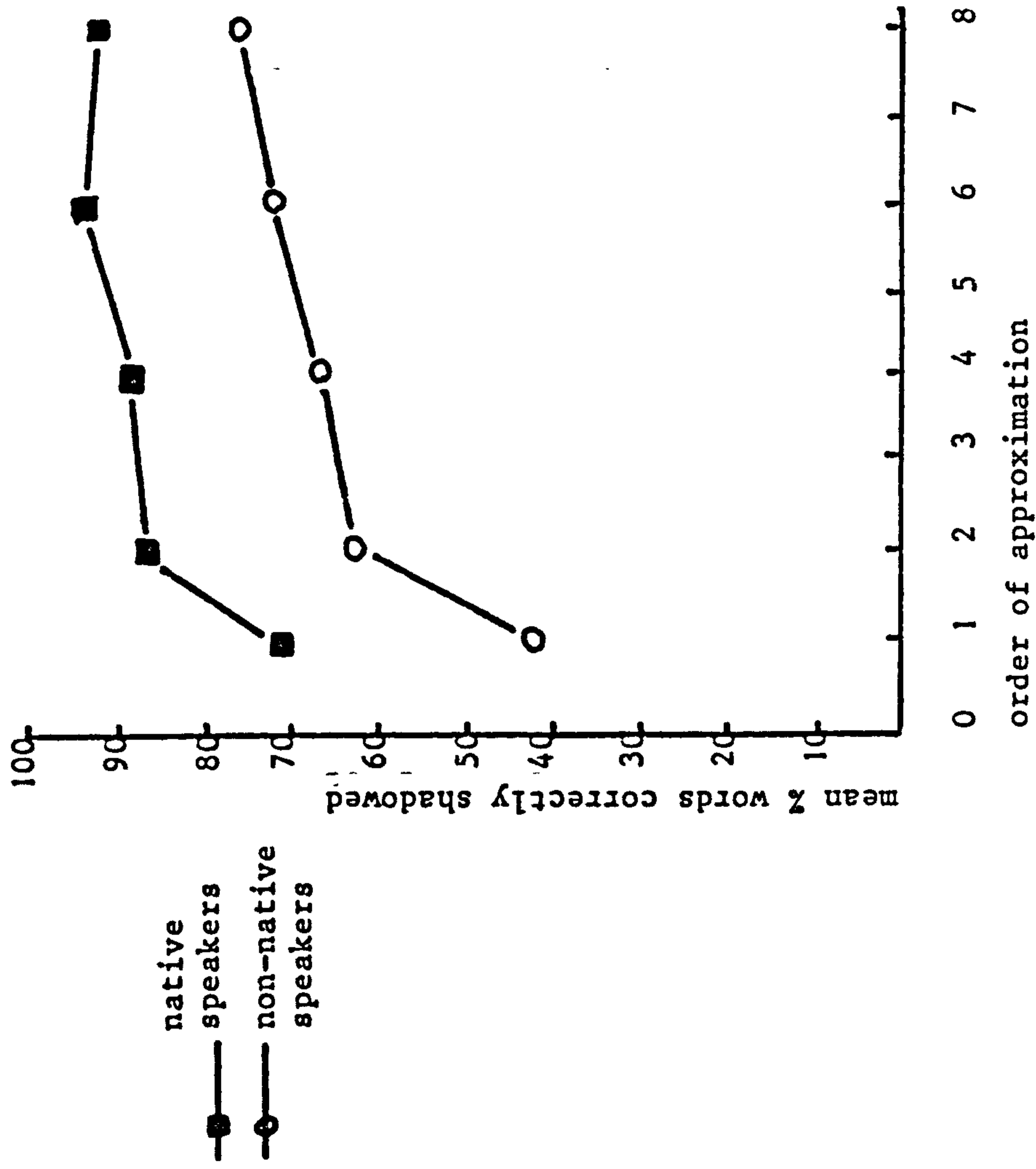
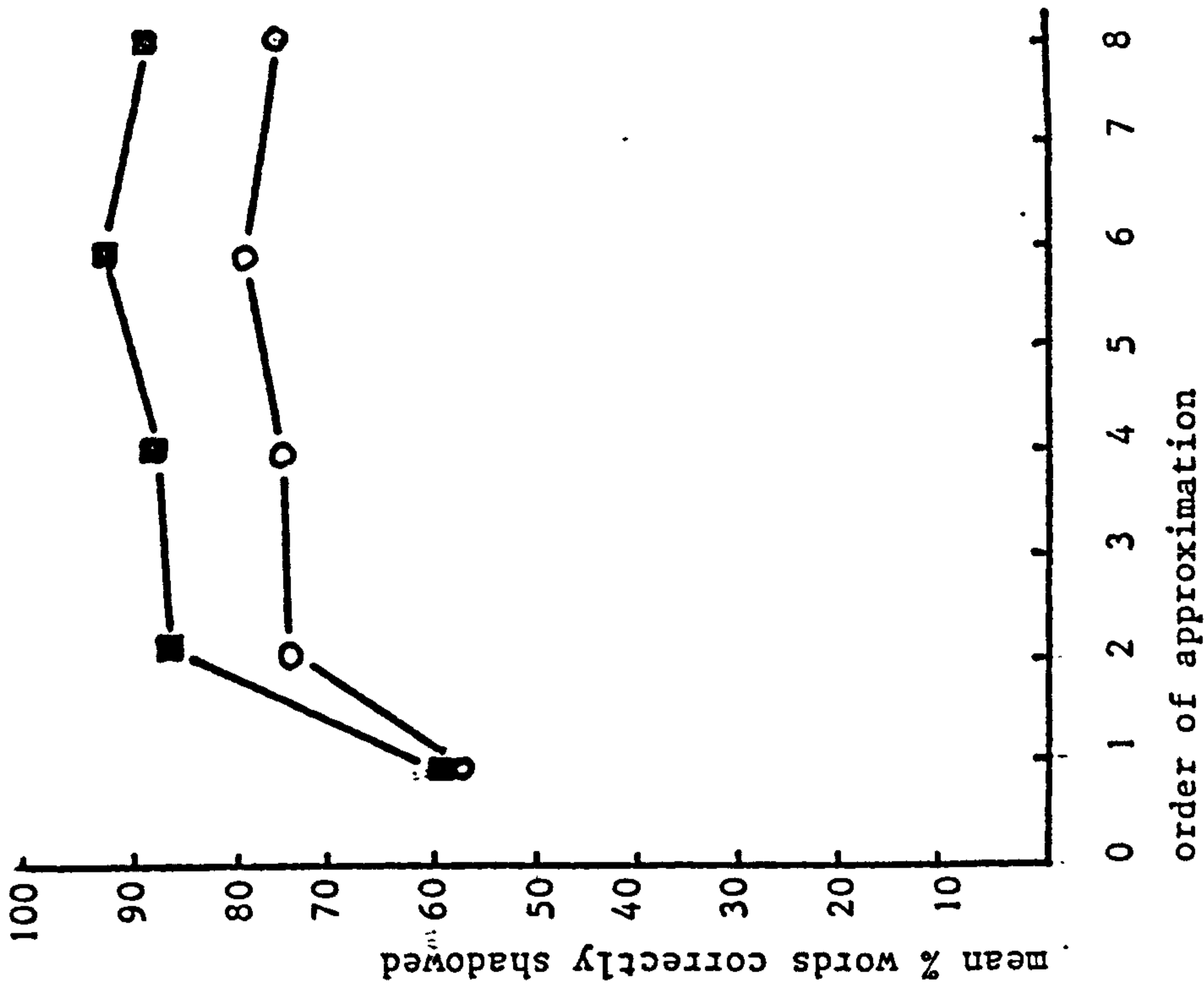


Figure 2-4b

French language passages



.2-5.        EXPERIMENT 1.

This section reports an experiment that is a straightforward replication of the original work by Miller and Selfridge, but which compares a group of native English speakers learning French with a group of native French speakers. The experiment is designed to test the hypothesis that foreign language learners do not make use of the redundancies that characterize higher orders of approximation.

## 2-5-1 Materials.

Two sets of materials were prepared for this experiment, consisting of strings of ten and twenty words at six different orders of approximation to French.

The 2nd, 4th, 5th and 6th orders of approximation were based on sample passages from Taylor and Moray (1960). The same passages were used by Triesman. This collection does not contain a 3rd order passage, so this order of approximation was omitted from the materials. (Taylor and Moray do not discuss why 3rd order approximations were not included in their listing.) Two pilot experiments suggested that Taylor and Moray's materials needed to be handled with some caution. Two problems emerged in these pilot studies. Firstly, Taylor and Moray's material contains some words of very low frequency, and it seems unreasonable to expect learners to be able to handle such words. In order to avoid this problem, each word in the strings selected for testing was checked for its frequency. Words not appearing in Gougenheim (1958) - a dictionary of the first 5,000 most frequent words in French - were replaced by a word of similar meaning and of the same form class (usually a superordinate term) which was listed by Gougenheim. Only a small number of such changes were necessary and these are recorded in Appendix 2A. It might be objected that this correction seriously interferes with the basic characteristics of the statistical approximations, but this objection does not seem to be an overriding one. The changes made do not in any way distort the syntactic patterns being studied, and furthermore, using more frequent rather than less frequent words should, if anything, reduce the differences between learners and native speakers, and to that extent the correction works against the

hypothesis, rather than in its favour. It is obvious that learners are going to behave differently if they do not understand the words that they are dealing with, but this is not an interesting problem; what we are concerned with is the differences that arise between native and non-native speakers even when the words used can reasonably be assumed to be known by the learners.

A second problem found during the pilot phase was that it is not clear from Taylor and Moray's brief discussion of their approximations how they intended elided forms to be handled. They appear to have counted forms like l'homme as one word, while la femme is counted as two. Other instances of this problem are qu'il, j'ai, c'est etc. This type of formative seems to be much more frequent in French than in English. It is not easy to see how these forms ought to be read or scored. Both l'homme and la femme are composed of a determiner and a noun, and it seems unreasonable to count the former as one word and the latter as two. This problem was avoided in the materials devised for the experiment by ensuring that no such sequence occurred in the sample strings. This unfortunately cut down the range of choice, and meant that only short strings of ten and twenty words could be used.

In addition to the higher order strings based on Taylor and Moray, a set of 1st order strings was formed by random sampling from a French novel, and a set of 0-order strings was formed by random sampling from Gougenheim (1958). Two further 0-order strings were also constructed in this way. These strings were of seven and thirteen words respectively; they were used as practice strings and were not scored.

A complete set of strings will be found in Appendix 2A. This appendix also records details of the alterations made on the grounds of frequency.

The strings were tape recorded by a phonetician at a rate of approximately one word per second. This is roughly the same speed as Miller and Selfridge's presentation. They recorded their lists "slowly and distinctly.....with a short pause between words." (p247). There are two main differences between the presentation used here and that of Miller and Selfridge. Whereas Miller and Selfridge's lists were read on a monotone, ours were read with each word given a low falling intonation. This seemed more natural than a monotone, but was not expected to influence the results in any substantial way. Secondly, where Miller and Selfridge presented all ten-word lists followed by all the twenty-word lists; in this experiment, the ten-word list and the twenty-word list at each order of approximation were presented one after the other, starting at 0-order and working upwards. The results of the native French speakers are sufficiently close to those reported by Miller and Selfridge for native English speakers to suggest that this change was not significant either.



## 2-5-2 Subjects.

Thirty subjects were tested in all. Fifteen of these, the learner group, were English school-children studying French at A-level. This group consisted of fourteen girls and one boy.<sup>5</sup> The second group consisted of fifteen native speakers of French. Twelve of these were students attending a summer course at Cambridge University. The other three were staff members of Cambridge and London Universities.

## 2-5-3 Method.

Subjects were instructed to listen carefully to the word lists on the tape, and, when each list ended, to write down in an answer booklet as many of the words as they could remember. The subjects were told that they should not expect to remember all the words. These instructions closely resemble those given by Miller and Selfridge. Instructions were given in English to the learner group, and in French to the native speaker group. Testing was done individually or in small groups, each subject listening to the materials through a pair of headphones. The test session lasted just over twenty minutes.

For scoring purposes, a word was counted correct if it was a clear attempt at one of the words in the original list. This is a lenient criterion, which tends to raise the scores of the learner group.

Homophones were marked correct - e.g. sept for cette, né for nez.

Also marked correct were verb forms which are frequently confused by

native speakers - e.g. donnais for donné. Spelling errors were ignored except where the error resulted in a different word, so collère and plouvoir for colère and pleuvoir are both counted as correct.

#### 2-5-4 Results.

The full detailed results, showing the scores of the individual subjects are given in Appendix 2B. A summary of these results is to be found in Table 2-1, and Table 2-2 shows the results of a three-way analysis of variance to which these scores were subjected. The main effects in this analysis were Group, String Length, and Order of Approximation.

As expected, all three main effects were significant. The native speaker group performed consistently better than the learners. A higher percentage of words is recalled on the shorter strings than the long ones. When the two groups are treated as a whole, the higher orders of approximation are recalled rather better than the lower orders. All these results are in line with previous work, and none is surprising. Two of the interactions are also significant, however, and these are discussed in detail below.

1: the interaction between length and order of approximation

Table 2-3 summarizes the results in terms of these two variables, combining the scores of both groups. The main source of the variance seems to be that the higher orders of approximation are recalled much better for short strings than they are for long ones.

Table 2-1.

Mean % words correctly recalled at different orders of approximation.

ORDER OF APPROXIMATION	0	1	2	3	4	5
10 words Native Speakers	50.0	62.0	85.3	98.0	98.7	93.3
Learners	38.7	54.0	50.0	82.0	64.7	59.3
20 words Native Speakers	34.3	40.3	72.3	66.7	69.3	71.0
Learners	26.3	31.3	43.0	33.0	28.7	28.7

Table 2-2.

Summary of Analysis of Variance.

SOURCE OF VARIANCE	DF	SS	MS	F
Subjects	29	70972.3		
Group	1	56876.7	56876.7	112.98**
Error	28	14095.6	503.4	
Length	1	52925.6	52925.6	314.04**
Length 1 group	1	370.1	370.1	2.20
Error	28	4718.9	168.5	
Order of Approximation	5	47525.6	9505.1	71.01**
Group 1 order	5	13297.8	2659.6	19.87**
Error	140	18741.1	133.9	
Length 1 order	5	9625.6	1925.1	16.67**
Length 1 group 1 order	5	1407.8	281.6	2.44*
Error	140	16164.4	115.5	
Residual error	330	164777.1		

\*\* =  $p < .001$

\* =  $p < .05$

Table 2-3

Mean % words recalled correctly for each test string.  
Scores for native speakers and learners combined and rounded to nearest whole percent.

ORDER OF APPROXIMATION	0	1	2	4	5	6
10-word strings	44	58	68	90	82	76
20-word strings	30	36	58	50	49	50

The difference between the two sets of scores is small for the lower orders, but much larger for the higher orders, though there is no clear relationship between the size of this difference and order of approximation. This interaction is not relevant to the main hypothesis which is concerned with the group factor, and it will not therefore be discussed further.

2: the interaction between group and order of approximation.

This interaction is the critical one as far as the hypothesis that we are examining is concerned. Broadly, it was argued that the native speaker and learner groups might be expected to react differently to increased contextual constraint in the material to be remembered. The fact that this interaction is significant lends some support to this view. Table 2-4 summarizes the scores of the two groups in terms of order of approximation, combining the scores for both lengths of string. The figures suggest that there is a marked increment for the native speaker group when they are asked to recall higher orders of approximation, but this increment is not apparent in the scores of the learner group. The exact nature of the interaction is not wholly clear from these scores, however, and in order to clarify what is involved, each set of scores is discussed separately below.

a: the twenty word strings.

These scores are summarized in Table 2-5 and Figure 2-5a.

The native speaker group show the typical pattern associated with recall of statistical approximations. There is a steady improvement in recall from 0- to 2nd-order, and then the scores level out. The

2nd-order score is actually rather higher than might have been expected, but the relatively small differences between the higher orders of approximation are not significant.

The scores of the learner group show a rather different picture. There is a slightly higher score on the 2nd-order approximation, but after this the scores tail off to a level that is almost identical to the 1st-order score. None of these scores is significantly different from the 1st-order score, and this finding seems to offer clear support for the claim that these learners are not responding to the syntactic structure that characterises the higher orders of approximation.

The rather high 2nd-order score, though not significantly different from the 1st-order score is slightly puzzling, however. The simplest explanation of this high score seems to be that some uncontrolled factor is at work; this would account for the rather high 2nd-order score found with the native speakers too. On the other hand, the relatively large standard deviation of the learners' 2nd-order scores suggests that this might be an over-simplification. Closer inspection of the raw scores suggests that while for the group as a whole the difference between 1st- and 2nd-orders of approximation is not significant, seven of the fifteen subjects scored higher on this 2nd-order string than on any other. If the group is split into two subgroups on this criterion, a picture emerges which is slightly less confusing. (see Table 2-6 and Figure 2-5b.)

Here subgroup A shows no improvement on higher orders of approximation. For this group, the difference between 1st- and 0-order is significant ( $t=2.706$ ,  $p<.05$ ) but the 1st-order score does not differ

Table 2-4.

Mean % words correctly recalled for each test string  
 Scores for long and short strings combined and rounded to the nearest  
 whole percent.

ORDER OF APPROXIMATION	0	1	2	4	5	6
Native Speakers	42	51	79	82	84	82
Learners	33	43	47	58	47	44

Table 2-5.

Means and Standard Deviations of Group Scores for the 20-word strings.

ORDER OF APPROXIMATION	0	1	2	4	5	6
Native Sepakers	6.8	8.1	14.4	15.3	13.9	14.2
S.D.	1.5	2.1	3.0	3.1	2.9	3.1
Learners	5.2	6.3	8.6	6.6	5.7	5.7
S.D.	1.4	1.9	2.7	1.3	1.8	1.6

Table 2-6.

Means and Standard Deviations for Sub-Groups A and B of the Learner  
 Group.

ORDER OF APPROXIMATION	0	1	2	4	5	6
Sub-group A	4.6	6.5	6.8	7.0	5.6	5.1
S.D.	0.9	2.1	1.1	1.4	1.7	1.4
Sub-group B	5.7	6.3	10.7	6.4	5.9	6.6
S.D.	1.8	1.3	2.3	1.1	1.8	1.4

Sub-group A comprises Subjects L1, L4, L5, L6, L7, L8, L12 and L15  
 Sub-group B includes Subjects L2, L3, L9, L10, L13 and L14.

Figure 2-5.

Mean percent correct words recalled. (20-word strings).

Figure 2-5a.

Native speakers ■  
Learners ○

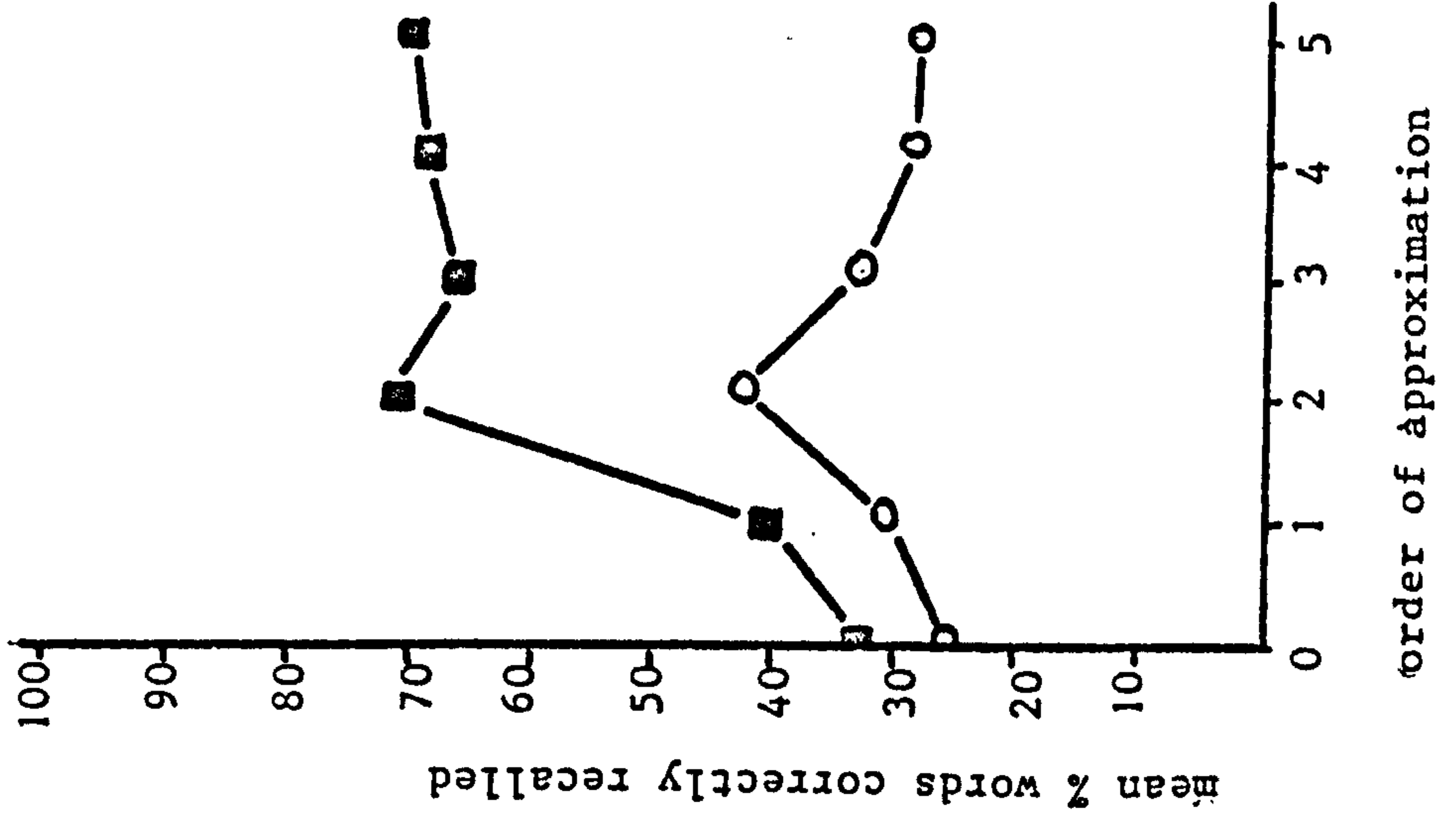
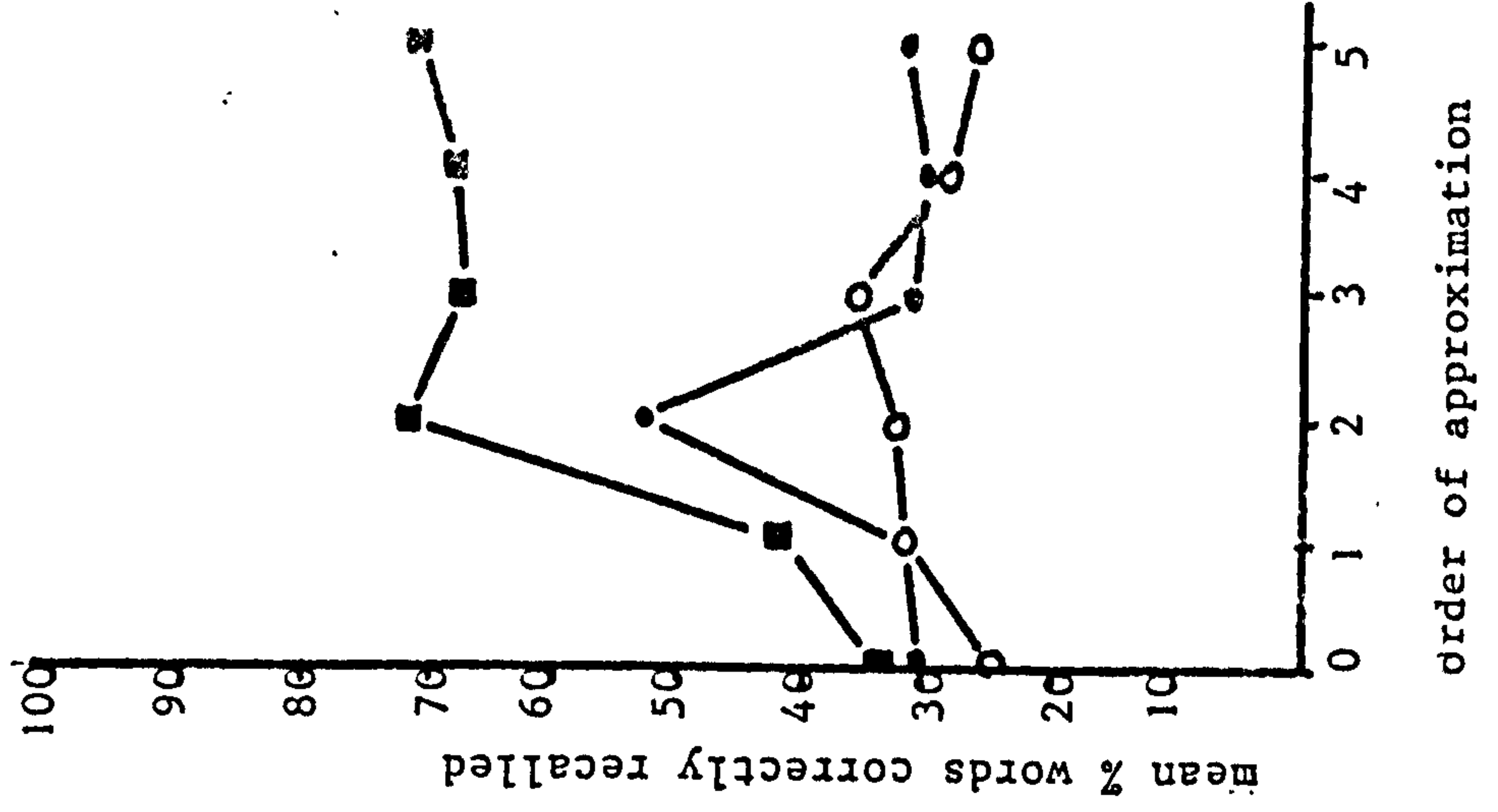


Figure 2-5b.

Native speakers ■  
Learners Sub-group A ○  
Learners Sub-group B ●

\*see text for discussion



significantly from any of the other scores. These results clearly support the word list conjecture. Subgroup B shows a significant difference between the 2nd-order score and the others, but the other scores are not significantly different from each other. One possible explanation of this finding is that this group of subjects is able to handle the short range constraints that characterize 2nd-order approxiamtions, but that they are not able to manage the larger constituents that are found in higher order strings. It is not clear, however, whether this is an adequate explanation, and for the moment this 2nd-order score must remain as a puzzle. There is no significant difference between the sub-groups except on this 2nd-order score.

b: the ten word strings.

These scores are summarized in Table 2-7 and in Figure 2-6a.

Once again, the native speakers show the typical statistical approximation curve, with the low orders significantly worse than the higher orders, and only minor differences between these latter. (The low score at 6th-order is almost entirely due to subjects substituting rouge for rose in this string.)

The learner group does not conform to this pattern, however. The 1st-order score is significantly higher than the 0-order score and significantly worse than the 4th-order score ( $t=4.78$ ,  $p<.001$  and  $t=4.4$ ,  $p<.01$  respectively) but the 1st-order score does not differ significantly from the other scores, all of which are surprisingly low. This lack of a significant difference is surprising, as the higher order strings do appear to be better recalled, but as in the case of the twenty-word strings, the standard deviations of the



Table 2-7.

Means and Standard Deviations of Group Scores. (10-word strings).

ORDER OF APPROXIMATION	0	1	2	4	5	6
Native Speakers	5.0	6.2	8.5	9.8	9.9	9.3
S.D.	1.4	0.9	1.4	0.4	0.3	0.5
Learners	3.9	5.5	5.0	8.2	6.5	5.9
S.D.	1.2	0.7	1.3	2.3	1.8	1.8

Table 2-8.

Mean words recalled by Learners. (10-word strings).

ORDER OF APPROXIMATION	0	1	2	4	5	6
L7 and L8	4.5	6.0	4.0	3.5	4.0	3.0
Remaining S's	3.7	5.4	5.2	8.9	6.8	6.4
S.D.	1.3	0.7	1.4	1.4	1.6	1.5

Table 2-9.

Mean Raw Scores for each length of string.

	Natives	Learners
20-words	11.8	6.4
10-words	8.1	5.8

Table 2-10.

Mean Raw Scores for each string length at each order of approximation. This table should be compared with Table 2-1 where the means are given in percentage terms.

ORDER OF APPROXIMATION	0	1	2	4	5	6
10-words Native Speakers	5.0	6.2	8.5	9.8	9.9	9.3
Learners	3.9	5.5	5.0	8.2	6.5	5.5
20-words Native Speakers	6.9	8.1	14.5	13.3	13.9	14.2
Learners	5.2	6.3	8.6	6.6	5.7	5.7

Figure 2-6.

Mean percent correct words recalled. Figure 2-6a  
(10-word strings).

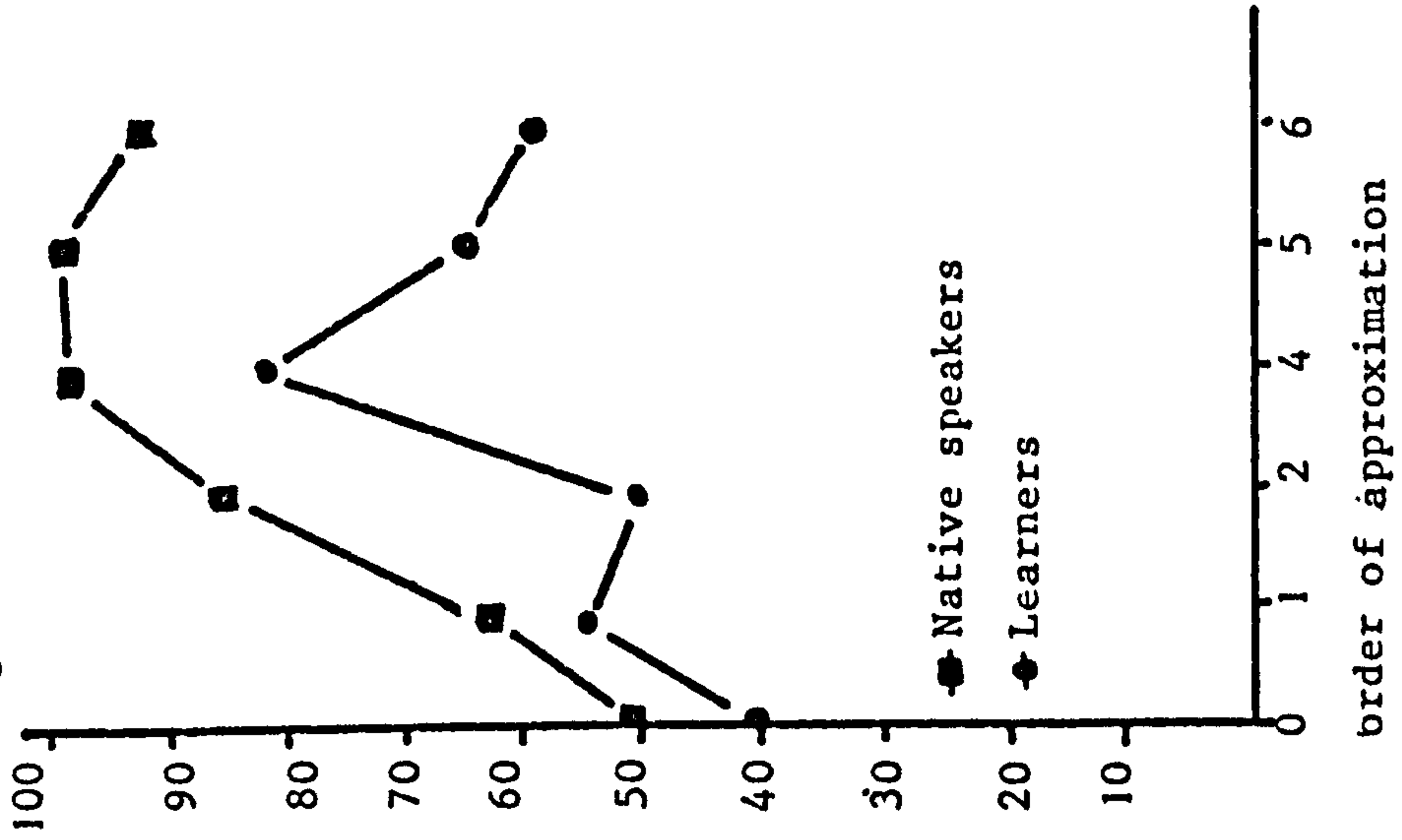
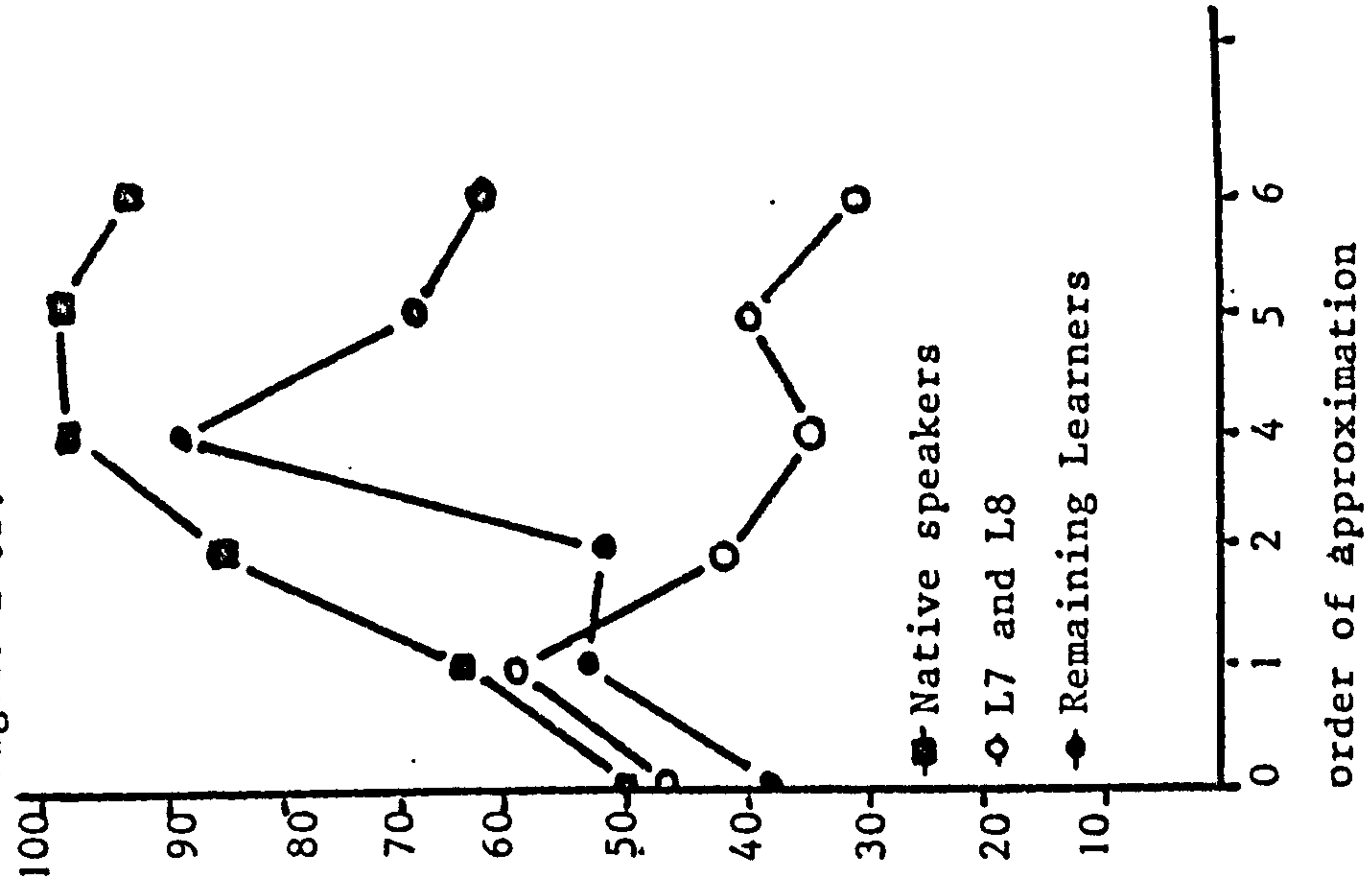


Figure 2-6b.



higher order scores are relatively large, and this again points to discrepancies in the performance of individual subjects as an important factor.

Two subjects, L7 and L8, did in fact show a steady decline in performance. Their scores, and the scores of the remainder of the group are shown separately in Table 2-8. Removing these two scores clearly reduces the variability, though the range of scores on higher orders of approximation in the remainder is still quite high compared to that of the native speaker group.

Leaving aside for the moment, the very high 4th-order score, the results suggest a small improvement on high orders of approximation over low ones. Though this improvement is not nearly so marked as for native speakers, the 5th- and 6th-order scores are significantly higher than scores on 1st-order approximation. ( $t=2.91$  and  $2.28$  respectively,  $p<.05$ )

The 4th-order score remains puzzling, however. It is not possible to ascribe this very high score to the subjects' being able to take advantage of short-term constraints. If this were the case the 2nd-order score should also be high, but it is in fact quite low. One possible explanation is that the string used in the experiment contains a large number of very short and frequent words, which may contribute to the high score. There is some evidence to suggest that length and frequency of words are more important for learners than for native speakers, though the data available here is insufficient to clarify this point. If we assume that the 4th-order score is artificially inflated, then the pattern that emerges is one that clearly contradicts the prediction made by the word list conjecture. The group as a whole are able to take some advantage out of the

Figure 2-7.

Mean number of words correctly recalled  
at each order of approximation

Figure 2-7a  
Native Speakers

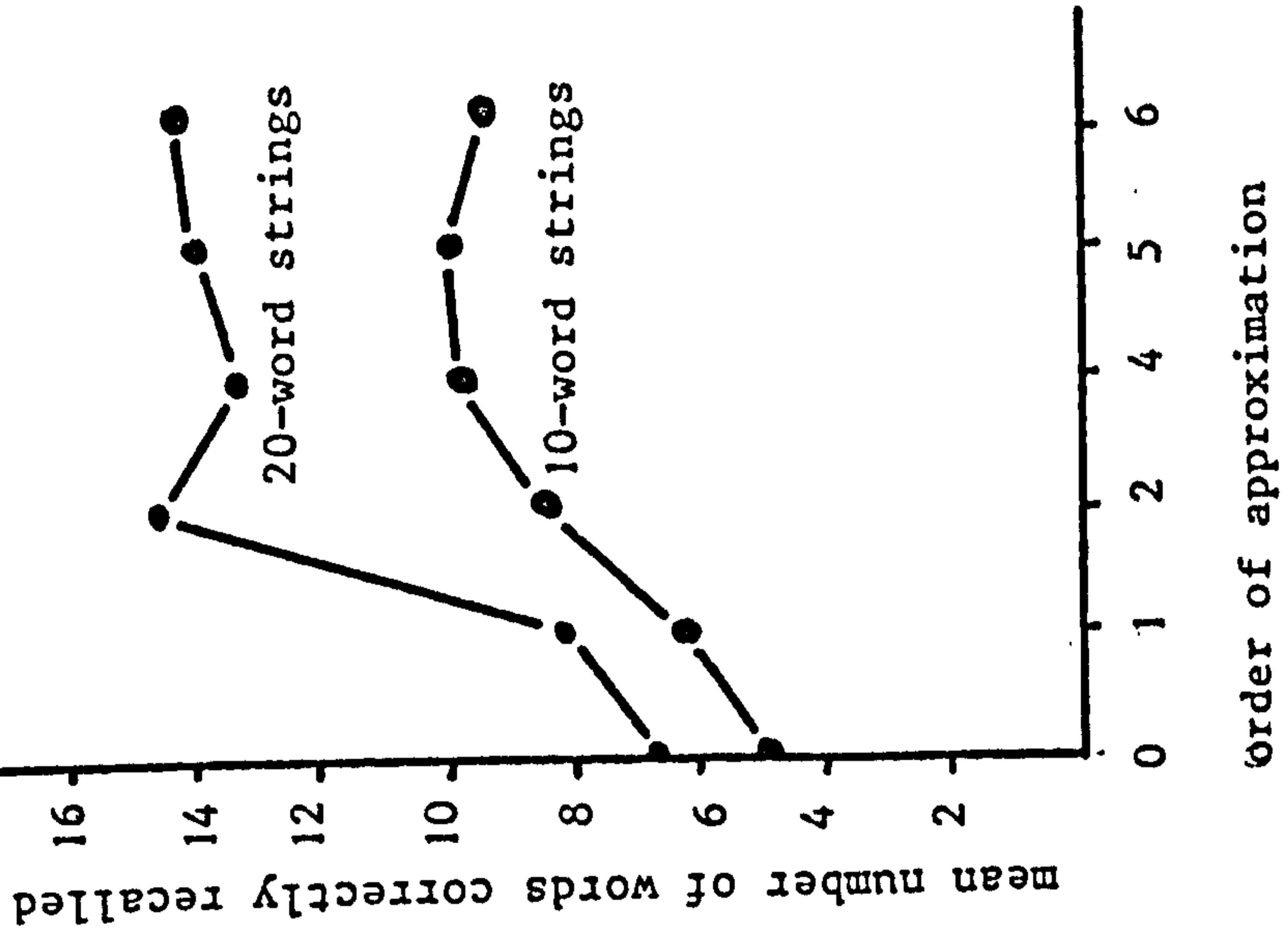
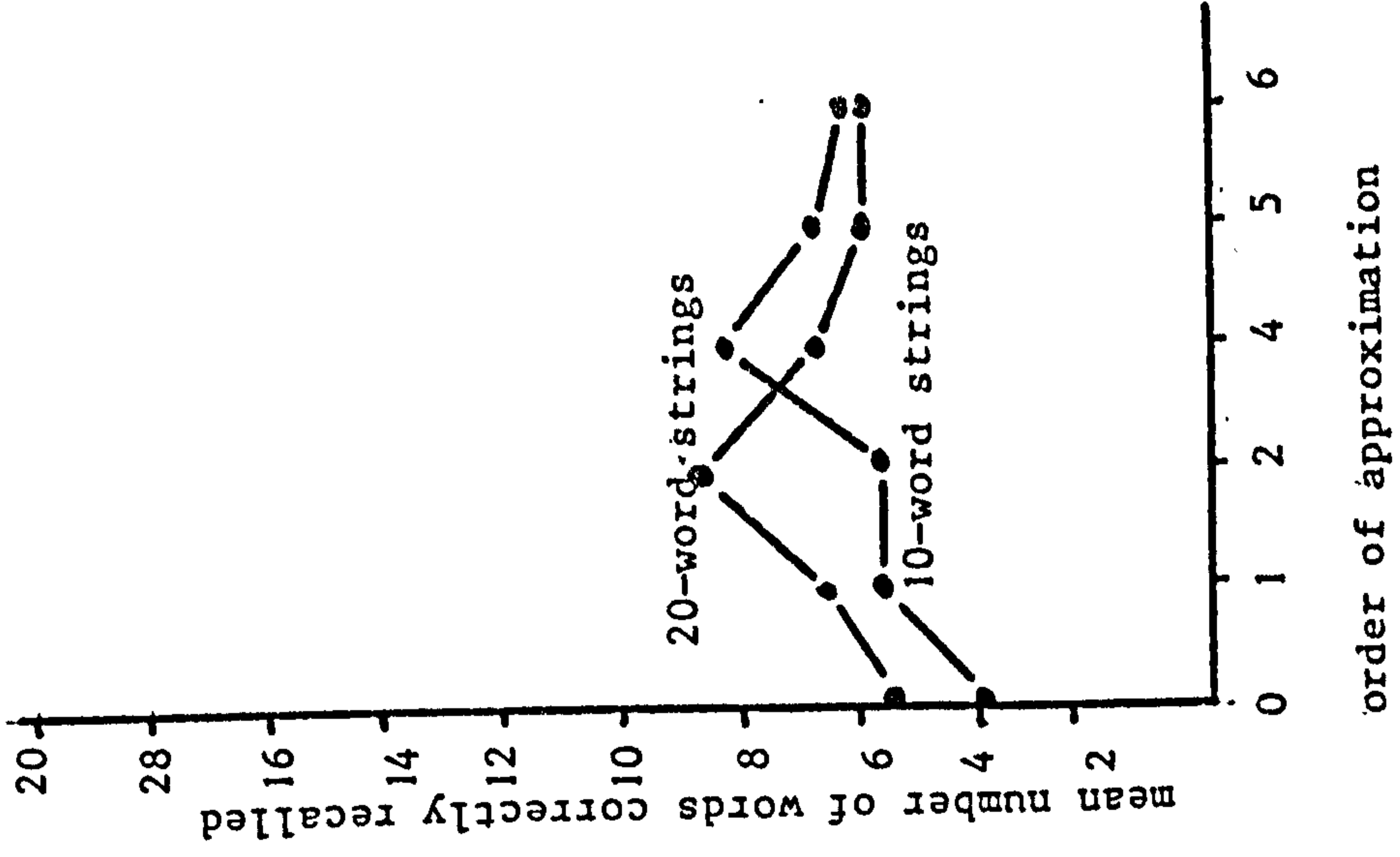


Figure 2-7b  
Learners



increased contextual constraint that is available in the higher order approximations. This advantage is not great, however, and amounts to an average increment of only one and a half words, far less than the improvement shown by the native speakers. Two of the subjects (L7 and L8) did fail to show any improvement on higher orders of approximation, suggesting that the word list conjecture may possibly be true for some learners. The numbers involved are so small, however, that it would be difficult to support this speculation with any degree of confidence.

The implications of these findings are discussed further in section 2-7.

3: the interaction between group and length. "

The analysis of variance summarized in table 2-2 showed that this interaction was not significant. This result may, however, be an artefact of the way in which the combined scores are handled. It is customary in reporting the results of experiments using statistical approximations to discuss the findings in terms of the percentage of words recalled, but this, of course, can obscure differences that might occur between the different lengths of string in terms of the absolute number of words recalled correctly.

Table 2-9 shows the mean raw scores for each string length, ignoring the differences between the several orders of approximation. The point to note about these figures is that while the native speakers improve their recall figure by a factor of 50% over the short strings, the learners absolute scores are almost identical on the 10 and 20 word strings. A second analysis of variance in which raw

scores rather than percent scores were used showed that this interaction is highly significant. ( $F = 46.1$ ,  $p < .001$ ).

Table 2-10 shows these summary scores expanded to show the mean raw scores on each order of approximation. Notice how narrow is the range of scores for the learners compared to that of the Native Speakers (cf. also Figure 2-7).

This surprise finding is not directly relevant to the argument in this chapter. It is clearly of some importance, however, and will be given further consideration in Section 2-7.

## 2-5-5 Experiment 1: Summary

The results reported in the previous sections suggest that the subjects do not all react in the same way towards increased contextual constraint. Subjects L7 and L8 show no evidence at all of using contextual factors to increase their memory span: syntactic structure appears to be wholly irrelevant to them. The numbers involved in this group are small however, and it is perhaps unwise to make too much of these results which clearly do support the Word-List Conjecture.

The remaining subjects are alike in that they show a very small improvement on high orders of approximation relative to low ones, as long as the strings to be remembered are short. This improvement is much less marked than that found for native speakers, and the observed increment amounts to less than two words on average. Though this finding contradicts the prediction made by the word list conjecture, it suggests that syntactic structure does not play a major part in the performance of non-native speakers; what syntactic effects are operating seem to be minimal, and there is clearly a major difference between native speakers and learners in this regard.

In the case of the longer twenty word strings, even this very small increment disappears, however. It seems that when subjects are faced with a message that exceeds their storage capacity by a very large amount, they do revert to a word by word processing strategy which is very inefficient.

A few of the subjects seem to be able to take advantage of the short range constraints that characterize the second order approximation to improve their score on this string. Here again, however, the improvement is small compared with that found for native speakers. It is possible that this high score is an artefact, as the native speaker group also scored unexpectedly highly on this string.

26.        EXPERIMENT 2.

Experiment 2 replicates the work reported in Experiment 1, but uses a more advanced group of learners as experimental subjects. The principal reason for studying subjects at this level was to find out whether syntactic effects of the type under investigation could be established for a more advanced group of subjects. A result of this type would indicate that the word list conjecture might apply only to a very restricted range of non-native speakers whose proficiency was severely limited. If, on the other hand, it could be shown that even fairly advanced students differ from native speakers in their response to syntactic structure, then it would be apparent that we are dealing with a phenomenon of some importance.



### 2-6-1 Materials, Method and Subjects.

The materials used and the method followed were identical to those used in Experiment 1.

The subjects were twelve students at Birkbeck College, London University, who were following the first year of a degree course in French.

### 2-6-2 Results.

The scores of the individual subjects will be found in Appendix 2B. Table 2-11 shows the mean scores of this group on each of the test strings, and for comparison those of the native speaker group used in Experiment 1. The scores of the learner group were subjected to a two way analysis of variance in which the main effects were length of string and order of approximation. This analysis is summarized in Table 2-12.

Both the main effects are highly significant, as also is the interaction between length and order of approximation. Table 2-13 shows the mean scores of the group on all the 10 and 20 word strings. The figures show that the short strings are recalled better than the long ones. Here again, however, measurement in terms of raw scores rather than percentage scores probably makes more sense. A further analysis on these lines shows that the length effect is still significant ( $F=82.2$ ,  $p<.001$ ) but that the twenty word strings are recalled slightly better than the 10 word strings. The difference between the two means amounts to only 1.4 words, however.

Table 2-11.

Mean % words correctly recalled at each order of approximation.

ORDER OF APPROXIMATION	0	1	2	4	5	6
10-words Native Speakers	50.0	62.0	85.3	98.0	98.7	93.3
Learners (L <sub>2</sub> )	47.5	47.5	61.7	94.2	85.8	82.5
20-words Native Speakers	34.3	40.3	72.3	66.7	69.3	71.0
Learners (L <sub>2</sub> )	29.2	30.8	53.3	51.3	49.2	39.2

Table 2-12..

Summary of analysis of variance.

SOURCE OF VARIANCE	DF	SS	MS	F
Subjects	11	5301.9		
Length	1	27639.1	27639.1	255.0 **
Error	11	1192.2		
Order of Approximation	5	24780.0	4956.0	35.1 **
Error	55	7759.5	141.1	
Length 1 order	5	6845.3	1369.1	13.5 **
Error	55	5560.9	101.0	
Residual Error	132	73777.1		

\*\* = p < .01

Table 2-13a

Mean % words correctly recalled for 10- and 20-word strings.

20-word strings	42.2
10-word strings	69.9

Table 2-13b.

Mean number of words correctly recalled for 10- and 20-word strings.

20-word strings	8.4
10-word strings	7.0

The significant order of approximation effect seems to be due to a higher level of recall on the higher order strings relative to the low ones. The interaction is due to the very large percentage increase found with the 10 word strings compared with the very much smaller one found for the 20 word sequences. (cf. Figure 2-8a.) Again, however, the percentage scores are confusing, and a reanalysis in terms of raw scores will be found in Table 2-13 and Figure 2-8b. The interaction between length and order of approximation here is still significant ( $F = 5.6, p < .001$ ) but now the interaction seems to be due to the rather higher scores found in the low orders of approximation for the 20 word strings. At higher orders of approximation there is scarcely any difference in these absolute scores.

### 2-6-3 Summary.

The results of Experiment 2 show that the more advanced group of learners studied here performed in a way which was not unlike that of the native speakers studied in Experiment 1. They all show small increments on high orders of approximation relative to low orders. These increments are, however, much less striking than those found with native speakers, and this result indicates that even these more advanced learners are still less able to take advantage of the structural redundancies found in the higher order strings than the native speakers are. These findings are complicated, however, by the very low scores on the longer strings. The learners failed to score very much higher on these strings than they did on the higher order 10 word strings, and this indicates that there is some complex interaction between structure and length whose nature is not at all clear.

Figure 2-8.

Figure 2-8a.  
Mean percent recall at different orders of approximation.

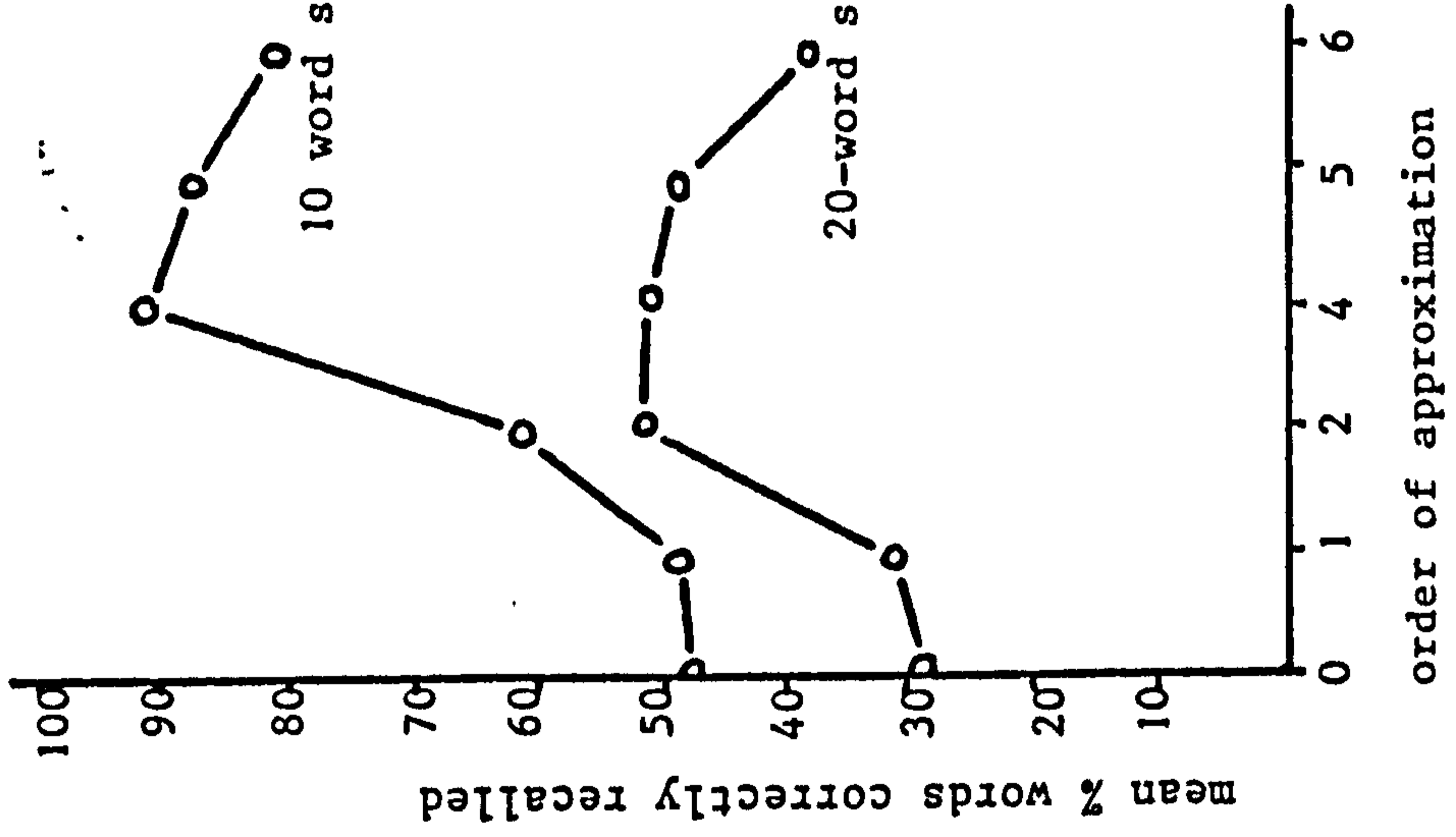
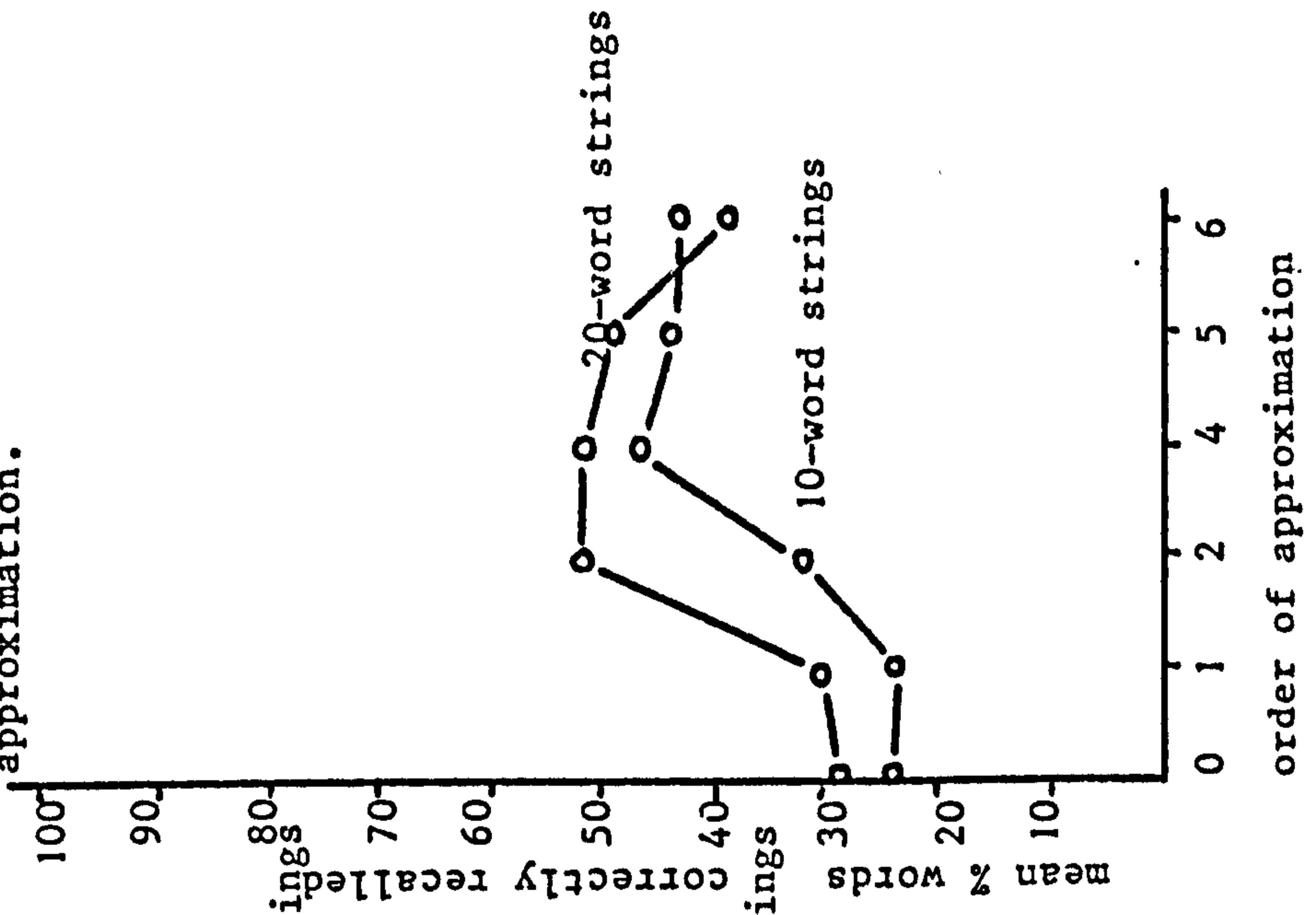


Figure 2-8b.

Mean number of words correctly recalled at different orders of approximation.



The main finding runs counter to the predictions made by the word list conjecture, but again it is apparent that there are some major differences between the learner group and the native speakers in their ability to make use of structural clues. The results seem to indicate that the word list conjecture in the form we have put it forward here does not hold for this group of subjects. As it stands it appears to be over-strong, and needs to be more carefully circumscribed.

## 2-7. DISCUSSION AND COMMENTS.

So far we have seen that there is a fairly broad quantitative difference between native speakers and learners in the recall of statistical approximations. Simply counting the number of items recalled is rather an insensitive measure, however, and masks a number of more subtle differences between the groups. This section will consider two further ways of scoring the results which will serve to highlight these differences: the sequence technique devised by Coleman (1963) and Deese and Kaufmann's (1957) suggestion that higher orders of approximation produce different strategies of recall in native speakers. These measures are applied only to the 20 word strings, as the 10 word strings are too short for the measures to be meaningful.

## 2-7-1 Coleman Scores.

Coleman (1963) criticised Miller and Selfridge's early work with statistical approximations on the grounds that it implied "that the only associations important to recall are those extending over sequences of less than six words." (p128)<sup>6</sup>. Coleman found it surprising that the standard curve for recall of statistical approximations should flatten out at all, and he argued that two main factors were responsible for this finding. The first factor, which will not concern us here, is the artificiality of the higher order approximations. The second factor Coleman distinguishes is the scoring method used by Miller and Selfridge. Miller and Selfridge scored simply the number of words recalled by the subjects without reference to either the order of recall or intrusions from words not in the original lists. Coleman argued that this method is not sufficiently sensitive to catch the differences between higher orders of approximation, and he suggested that as well as scoring single words correct, 2-word sequences, 3-word sequences, 4-word sequences correct and so forth should also be considered. Coleman's results showed that, scored in this way, higher order approximations do produce better recall than lower ones, and the longer the scoring sequence, the more positively accelerated the recall curve becomes.

While Coleman's analysis of his figures seems reasonable, it actually needs to be treated with some caution. Though he used six orders of approximation, the subjects were asked to recall only three passages, and the figures produced in Coleman's graphs represent not real scores, but amalgams. The figure for  $1\frac{1}{2}$  order of approximation, for example, is a combined figure derived from the mean score of one

Table 2-14.

Mean number of sequences of a given length at different orders of approximation. (20-word strings.)

ORDERS OF APPROXIMATION	0	1	2	4	5	6
2-word sequences						
Native Speakers	2.33	2.26	10.13	9.46	10.60	10.93
Learners (L2)	1.16	1.25	6.42	5.50	6.16	4.50
Learners (L1)	0.46	1.20	4.26	2.86	1.93	1.66
3-word sequences						
Native Speakers	1.00	0.80	7.73	6.73	3.41	8.53
Learners (L2)	0.41	0.25	3.92	3.41	4.00	2.41
Learners (L1)	0.06	0.02	2.33	1.53	1.06	0.60
4-word sequences						
Native Speakers	0.46	0.80	7.73	4.93	8.46	8.53
Learners (L2)	0.25	0.00	2.58	2.25	2.75	0.91
Learners (L1)	0.00	0.00	1.20	0.80	0.60	0.06
5-word sequences						
Native Speakers	0.26	0.00	4.20	3.60	5.40	5.06
Learners (L2)	0.00	0.00	1.50	1.58	1.83	0.25
Learners (L1)	0.00	0.00	0.40	0.40	0.33	0.00
6-word sequences						
Native Speakers	0.13	0.00	3.06	2.66	4.26	4.00
Learners (L2)	0.00	0.00	0.91	1.00	1.08	0.00
Learners (L1)	0.00	0.00	0.13	0.26	0.13	0.00
7-word sequences						
Native Speakers	0.06	0.00	2.06	2.13	3.20	3.06
Learners (L2)	0.00	0.00	0.50	0.05	0.58	0.00
Learners (L1)	0.00	0.00	0.00	0.02	0.06	0.00
8-word sequences						
Native Speakers	0.00	0.00	1.13	1.66	2.20	2.20
Learners (L2)	0.00	0.00	1.66	0.16	0.33	0.00
Learners (L1)	0.00	0.00	0.00	0.00	0.00	0.00



group of subjects on a 1st-order string and a second group of subjects asked to recall a 2nd-order approximation. This procedure seems to be highly questionable, but Coleman does not attempt to justify it. The main effect seems to be that it smoothes out variations in the curve, but it should be borne in mind that the curves found in Coleman's report are based on only three data points, and not on six. It should also be pointed out that Coleman does not use a 0-order approximation. This is a pity, as one might expect some quite large differences between 0-order and 1st order approximations, and it would have been interesting to have data for sequences in recall of unconnected items.

Table 2-14 gives the sequence scores of the three groups studied by experiments 1 and 2. The scores are not conflated as in Coleman's paper, so the resulting curves are more variable and less linear. Nevertheless a picture emerges that is broadly compatible with that drawn by Coleman.

Inspection of these scores shows that the native speakers perform in much the same way as the native English speakers investigated by Coleman. The native French speakers show a very marked rise in the presence of long sequences on 2nd-order strings and higher ones, and though there is some fluctuation in the scores for the longer sequences, the scores for sequences of 2, 3, 4 and 5 words show a high degree of stability.

With the two learner groups, however, a rather different picture emerges. Both groups show evidence of long runs on the higher orders of approximation, but these sequences are much less frequent than those of the native speakers, and there are only sporadic

instances of runs longer than seven words. The less advanced group seem to be able to handle two-word sequences reasonably well, but sequences of more than four words are rare. Generally speaking, the longer the sequence, the larger is the difference between the native speaker group and the two learner groups.

These figures seem broadly to confirm the analysis of the standard scoring put forward above. The learner groups do show some improvement with higher order strings, but this improvement appears to be strictly limited. In the case of the less advanced group, there are strong indications that they are unable to process sequences of more than three or four words as a whole. The more advanced group do rather better than this, but even this group rarely exceeds 50% of the native speaker score, and as the sequence length increases, the native speaker scores exceed the learner scores by factors of three or four.

## 2-7-2 Serial effects in recall.

Deese and Kaufman (1957) argued that the higher scores found in higher orders of approximation were due in large part to the type of strategy used by the subjects to reproduce the string to be recalled. Deese and Kaufman showed that with low orders of approximation the probability of an item being recalled depends mainly on its position in the original string, and this produced the usual serial position effects, early and late items being recalled particularly well. Deese and Kaufman showed that with higher orders of approximation, however, subjects tended to adopt different recall strategies, and to recall items in the same order as they were presented. Deese and Kaufman demonstrate this claim by calculating the mean serial position in recall and correlating this figure with order of input. For low orders of approximation this correlation is not significant, whereas high orders of approximation and text produce significant correlations.

There are, however, some problems with this finding. Deese and Kaufman do not define what they mean by mean order of recall, and both the obvious ways of calculating this figure give rise to anomalies which make the results difficult to interpret.

Suppose, for example, that we calculate the order of recall for each subject and rank each of the input words accordingly, total the ranks for each word and divide by the number of S's recalling the word: i.e. we use formula 1:

$$\text{formula 1: } P = \frac{\sum R}{N}$$

where P = mean recall position  
R = recall position for each S  
N = number of S's recalling the word.

Now consider a word from the middle of a list which is recalled by

only one subject, in first position. Such an item would have a P score of 1, and would emerge higher than an item recalled in first position by 19 subjects, but in second position by the 20th, thus scoring a P of slightly less than 1. This outcome is clearly unsatisfactory.

The second alternative is to make N in formula 1 equal the number of S's taking part in the experiment, irrespective of whether they manage to recall the word in question or not. This too causes some difficulties. Words which are recalled by many S's automatically have a higher mean score than those recalled by few. Now consider a word which lies sixth in the input list. Whenever it is recalled it is recalled in 6th position, but it is recalled by only a small number of S's. The mean recall position of this word according to the revised formula 1 will be lower than 6, and the important fact that the word is always recalled in correct sequence if it is recalled at all is lost.

The main difficulty with the learner scores is that the gaps in recall are so numerous that it is difficult to find an appropriate statistical summary. One way round this difficulty is to correlate the input and output orders of each subject separately, and to ignore the gaps. A subject who recalls, say, half a string, basically in the correct order, but with a number of gaps would still have a high correlation between input and output orders. The problem can then be reduced to the proportion of subjects scoring such high correlations at the different orders of approximation. The results of such an analysis are shown in Table 2-15 and also in Figure 2-9.

Table 2-15.

Numbers of Subjects showing correlations of +1.00 between input order and output order at each order of approximations

ORDER OF APPROXIMATION	0	1	2	4	5	6
Native Speakers (N=15)	2	0	6	8	10	8
Learners (L2, N=12)	1	1	3	2	7	7
Learners (L1, N=15)	0	0	1	2	4	2

Figure 2-9.

Percentage of each group showing correlations of +1.00 between input and output at each order of approximation.

Figure 2-9a.  
Native Speakers. N=15

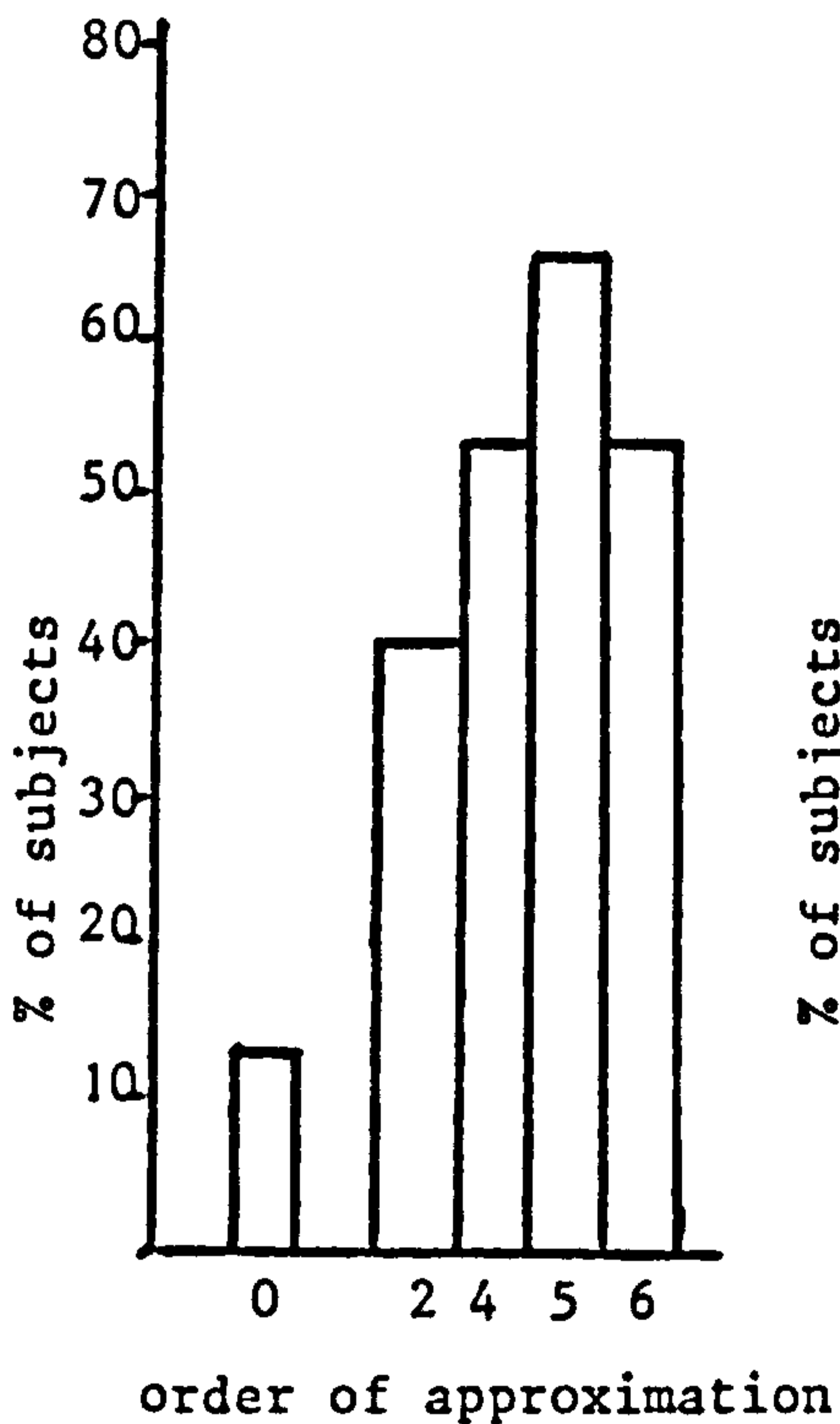


Figure 2-9b.  
Learners (L2). N=12

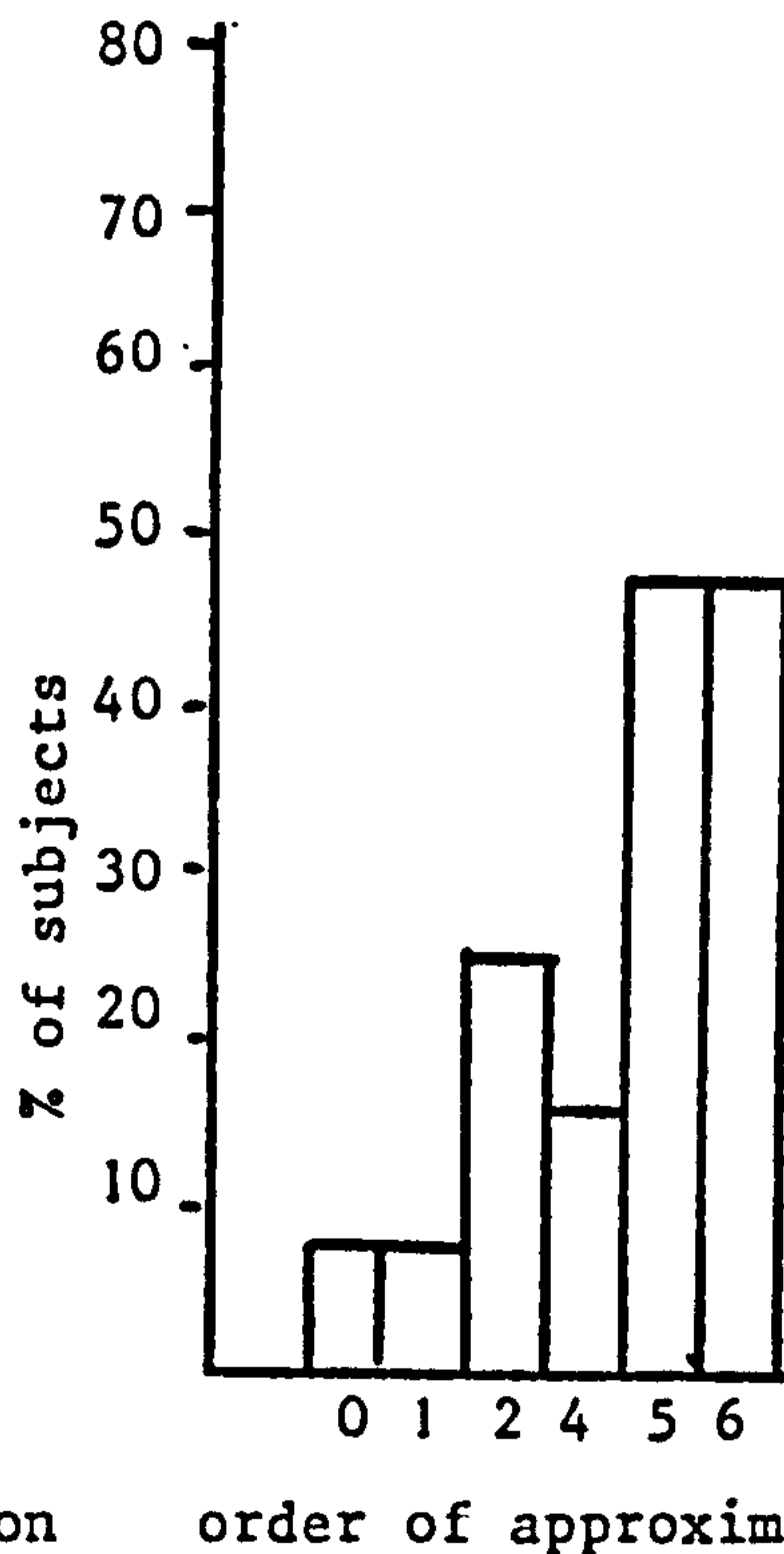
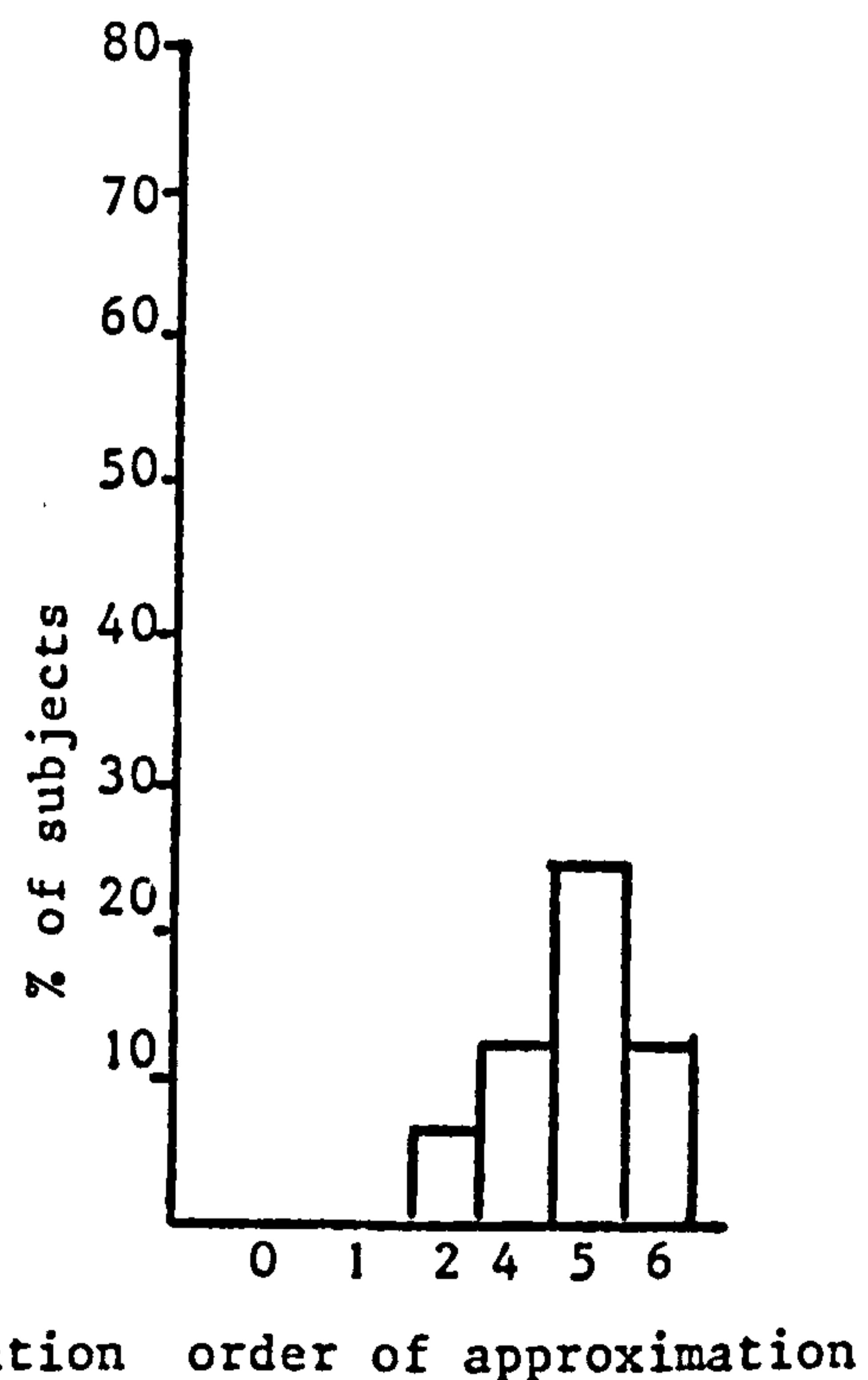


Figure 2-9c.  
Learners (L1). N=15



These data show that all the groups do better on the higher orders, though in the case of the weaker learner group, only a small number of subjects manage to perform at the criterion level, while a much larger proportion of the native speakers perform in this way. These figures should not be given too much importance, however, as the correlation measure is not really so reliable as it seems at first sight. There are two main reasons for this. Firstly, the test used penalizes a large number of small misorderings much less heavily than one large misplacement. Consider, for example, the following hypothetical cases:

Input order	Output 1	Output 2
1	2	8
2	1	1
3	4	2
4	3	3
5	6	4
6	5	5
7	8	6
8	7	7

Input order and Output 1 are highly correlated, while Input order and Output 2 are not ( $\rho = .904$  and  $.333$  respectively). Intuitively, however, the recall pattern in case 2 is "better" than that of case 1. It can be described simply in terms of one long sequence and a single word out of order, whereas case 1 is much harder to describe in simple terms like this. Case 2 seems to be a more natural misordering than case 1 where the elements you are dealing with are structured sequences of words.

A second problem is that the correlation on its own does not take into account the number of words used in each sequence. The weaker

learner group, for instance, have a total of 10 correlations of 1.00, but two of these contained only four items, and only one reached as many as 10. The native speaker scores average 14.75 words, and range from 11-19 - a much more impressive feat.

What this means is that figures shown in table 2-15 seriously under-rate the performance of the native speakers, and overrate that of the learners by not taking into account the length of string involved.

## 2-8. CONCLUSIONS.

In this chapter we have used statistical approximations in an immediate recall task in order to test the word list conjecture - a claim that non-native speakers are insensitive to syntactic structure in foreign language strings. The results of the two experiments reported here do not wholly support the conjecture. Only when less advanced learners are asked to handle long strings is there any clear evidence for a complete absence of a structural effect. On the other hand, the structural effects that have been found are extremely small - far smaller than those found in the native speakers tested - and this indicates that there are some major differences between native speakers and learners of a language in their ability to handle syntactically structured strings. It might therefore be possible to save the word list conjecture by modifying it so that it allows a very small degree of structuring behaviour to take place - for example by allowing some specific structure types to be handled easily, but not others. This line of argument will be pursued in Chapter 4.

Before moving in this direction, however, one further experiment which probes the word-list conjecture in its strong form will be reported. There are a number of reasons for choosing to follow this course of action instead of embarking immediately on tests of more circumscribed versions of the word list conjecture. These derive mainly from some criticisms that might be legitimately levelled at the experiments reported in this chapter.



Firstly it might be argued that Statistical Approximations are not natural language phenomena, and that they introduce distortions that invalidate any conclusions that one could draw about natural language processing. Statistical Approximations quite deliberately distort the normal expectations that listeners or readers have, consisting as they do of strings that are only partially interpretable. Furthermore, in the experiments reported here, this abnormality is reinforced by the use of a very slow presentation rate which removed even the usual clues to structure associated with rhythm and intonation, and artificially isolates and emphasizes the individual words. This is an important objection whose force is fully recognized here, and the experiments reported in the subsequent chapters all make use of more natural stimuli and less artificial tasks.

Secondly, and more importantly, it might be argued that immediate recall of long word strings is too crude a task to provide any clear indications as to what the real differences between the native and non-native speaker groups might be, so that although the results falsify the word-list conjecture, it is not immediately obvious how it might be revised in order to produce a more robust claim. Recall of long strings is a fairly complex task, and it is not clear where exactly in the processing of statistical approximations the differences between native speakers and learners arise. We have assumed that the differences reported in this chapter are due essentially to differences in the early stages of processing, but there is not really any compelling justification for this assumption. It is at least plausible that some of the differences might be due not to early processing, but to high-level output constraints that affect learners but not native speakers - say an inability to use syntactic

redundancies to facilitate output, rather than an inability to make use of syntactic redundancies to facilitate the input and storage of word strings.

There seems, therefore, to be a strong case for investigating at least one other global structure effect, before moving on to investigate the effects of particular structures on learners' performance, and by preference a structural effect that is produced with an experimental technique that makes few demands on the subjects at the output stage, by requiring only the simplest of responses. It was hoped that a structure sensitive task of this sort would enable us to confirm that there are some major differences between native and non-native speakers in their ability to use syntactic structure in the processing of sentences, and not just in the structuring of their responses. It was further hoped that this second experiment might provide some useful pointers about how the word list conjecture could be profitably refined.

CHAPTER 3. CLICK LOCATION.

3-1 Introduction

3-2 The Click Placement Paradigm

3-3 Click Placement and Foreign Language Learners

3-4 Experiment

3-5 Discussion

### 3-1. INTRODUCTION.

The experiments reported in Chapter 2 were concerned with a general syntactic effect rather than with the effects of any particular syntactic structure. This line of argument is pursued further in this chapter, which looks at the ability of foreign language learners to locate short bursts of noise (clicks) superimposed on sentences and lists of words.

There are two main reasons for choosing to investigate click location. Firstly, this type of experiment involves only minimal responses on the part of the subject. The experiments in Chapter 2 required the subjects to remember and reproduce long strings of words, both complicated tasks, and it is not easy to determine whether the differences found between native speaker and learner groups are due to input effects, output effects, or a combination of both. In contrast, the click placement task is very simple, and does not require the subject to make any responses in the foreign language. This should allow us, in principle at least, to establish whether the effects found in Chapter 2 are genuine perceptual effects, or whether the differences emerge at some later stage in the processing of the stimuli. The second reason for using click placement is that the general effect of structure (see below for a more detailed discussion) is so large that it makes it possible to use relatively natural stimulus material. Given the artificiality of the material used in Chapter 2, it seemed important to establish whether similar results were to be found using real language data.

### 3-2. THE CLICK PLACEMENT PARADIGM

#### 3-2-1 Ladefoged and Broadbent's Experiments.

The original click experiments were carried out by Ladefoged and Broadbent (1960). This paper reports a number of experiments, of which the first two will be considered here in some detail.

Ladefoged and Broadbent's first experiment compared two groups of subjects who were asked to listen to a series of short sentences that made up a simple story. Superimposed on each of these sentences was a click (a 200m.sec. capacitor discharge) and the subjects' task was to locate the objective position of this click as accurately as possible. One of the groups was given prior exposure to the passage, and marked the position of the click on a printed script. The second group had no prior experience of the materials, and were required to write down the word or words where they thought the click fell.

Ladefoged and Broadbent divided their scripts into a series of "slots", each slot consisting of either a word or the space between two words, and scored errors as follows; a click that was accurately marked scored zero; a click moved one slot leftwards (i.e. towards the beginning of the sentence) scored +1; a click moved two slots leftwards scored +2 and so on. Clicks that were moved rightwards (i.e. towards the end of the sentence) were scored in a similar fashion, but were given negative values. (Ladefoged and Broadbent do not comment on this apparently arbitrary decision.)<sup>7</sup>

No significant differences between the groups were found, but both groups were very inaccurate in their placements. Clicks tended to be reported as occurring in the word that preceded their objective

location - the mean scores for the two groups being 2.3 and 1.83 slots respectively.

Ladefoged and Broadbent's second experiment was a partial replication of the first, using the same sentences and task. In this experiment, all the subjects were given prior exposure to the test material, but in addition to the sentence task, the subjects were also required to locate clicks superimposed upon a sequence of digits. The subjects' responses were scored in the same way as in experiment 1. Ladefoged and Broadbent reported a large difference between the two sets of materials, in that though accuracy on the digit sequences was still low, with a mean displacement score of 0.7 per subject, these reports were considerably more accurate than reports of clicks in sentences, where the mean error rate is 2.0. Ladefoged and Broadbent report that all subjects did better on digit sequence than on sentences, and conclude that

"there seems to be something about an ordinary sentence which makes it harder to judge the location of an item superimposed on it." (p164)

Ladefoged and Broadbent offered two explanations of the fact that clicks were more accurately placed in digits than in sentences. The first explanation is couched in terms of an interaction between attention and the information carried by the strings. They point out that the leftward movement of the click is reminiscent of earlier experiments on "prior entry", where, for instance, a subject is required to watch a pointer move round a dial, and record its position when a bell sounds. Typically, subjects mis-estimate the position of the pointer, recording an earlier position, but this misplacement occurs only if the subject was expecting to hear the

bell sound. Ladefoged and Broadbent go on to argue that the amount of information carried by each item in the sequence of digits is relatively high - 3.1 bits per item - while the information conveyed by the items in the word lists is lower than this. Subjects are known to pay more attention when there is a high probability of an information bearing stimulus appearing, and this leads to a shorter reaction time.

"Consequently, when decoding the information in the sentences in the story, the subject does not pay as much attention to the words in the utterance as does when decoding the information in an utterance consisting of a series of digits. In the latter case his predisposition which is normally towards the superimposed item will be reduced: and therefore it is reasonable that he will be more accurate when the speech material consists of a series of digits." (p168).

This explanation of the click displacement in terms of information rate is not wholly satisfactory, as Ladefoged and Broadbent are aware.<sup>8</sup> Subsequent research using the click methodology failed to take up the idea, and much more influential was a second suggestion advanced by Ladefoged and Broadbent that click displacement might reflect the size of the unit of processing in speech perception. Ladefoged and Broadbent do not discuss this idea in any detail; rather it appears in the final paragraph, almost as an afterthought:

"We may conclude from this that the process of decoding information when listening to speech may involve operating on units which are somewhat larger than the duration of a single speech sound. Admittedly it is not possible to say much more about the units, for there is no definite maximum error in the location of superimposed items." (p169)

It is quite clear here that Ladefoged and Broadbent envisage their units as relatively small. Subsequent work, however, interpreted the suggestion rather more loosely, suggesting that units even larger than the word might be responsible for click displacements. Fodor, Bever and Garrett (1974) who were responsible for much of this work, summarize Ladefoged and Broadbent's suggestion as follows;

"(Ladefoged and Broadbent) interpreted this finding as related to the subject's perceptual segmentation of the stimulus strings. The suggestion is that clicks are displaced from their objective positions to the boundaries of perceptual units, and that the character of such displacements can be employed to estimate the size of these units. This interpretation takes click displacement to be a species of the well-known phenomenon of perceptual 'closure' - the tendency of perceptual units to resist interruption:" (p330)

Explanations of this sort, rather than the more limited interpretation of Ladefoged and Broadbent made the click paradigm an important research tool in the years that followed.



### 3-2-2 Development of the Click Paradigm.

Subsequent research using the click methodology showed that the displaced clicks were not moved in a random fashion. Rather there was a tendency for displaced clicks to move towards or into major constituent boundaries, and a considerable research effort was put into an attempt to show that the unit of perception corresponded in some way with units that could be easily identified in terms of the constituent structure of the sentences being tested. Despite its initial promise, this work seems to have rapidly become very limited in scope, concerned mainly with the claim that major clause boundaries in general, and the boundaries of deep structure sentoids in particular, were strong attractors of clicks, and therefore prime candidates for the boundaries of perceptual units. Some of this work shows considerable technical sophistication, but this often seems to have obscured basic questions rather than to have clarified them.

As an example of this tendency, consider a much quoted paper by Garrett, Bever and Fodor (1966). This paper attempted to rule out the possibility that acoustic correlates of constituent structure, such as intonation contours or pausing, rather than the constituent structure itself was responsible for the movement of clicks towards clause boundaries. In order to control for this possibility, Fodor, Bever and Garrett constructed a set of six pairs of sentences where each member of the pair contained a sequence of words in common, but differentiated by their constituent structure. An example is

provided below:

A) In her hope of marrying Anna <sup>\*</sup>was surely impractical

B) Your hope of marrying Anna <sup>\*</sup>was surely impractical

Here the underlined portions are common to both sentences, but in sentence A there is a major constituent boundary before Anna, while in sentence B, the major constituent boundary falls after Anna, which is the object of the verb in the first clause. Stimuli were prepared by making tape recordings of these sentences, and splicing the common portions of the sentences onto the different initial segments, thus making a pair of strings in which the final segments are acoustically identical, but syntactically distinct. Clicks were superimposed on this material, in one of the two positions marked by asterisks. In sentence A, the click over Anna occurs after the major clause boundary, and should therefore migrate leftwards. In sentence B, the click over Anna occurs before the major clause boundary, and should therefore migrate rightwards. Fodor, Bever and Garrett report that

"differences in responses to identical strings were predicted by the points of variation in constituent structure." (p30)

This interpretation was widely accepted at the time, and rapidly became part of the psycholinguistic folk-lore. There are, however, a number of difficulties with the findings. Firstly, Fodor, Bever and Garrett's main analysis concerns only clicks which migrate to the position either immediately right or immediately left of the points of variation. In the case of the sentences quoted above, this means that clicks appearing in the spaces surrounding Anna are counted. Four out of the six sentences tested show an asymmetry in that when these two positions are compared, clicks tend to prefer

the clause boundary to the other position. However, this is a very crude measure, as it ignores the very large number of clicks that do not migrate to either of these positions. On average, these two slots account for only 28% of the total responses, and in one case this figure sinks as low as 15%. It is difficult to see how evidence like this could be considered strong support for the claim that clicks tend to migrate into major clause boundaries.

Secondly, it is in any case doubtful whether the splicing technique is an adequate control of the acoustic correlates that Garrett, Bever and Fodor wish to eliminate. Consider, for example, the intonation of the strings. Garrett, Bever and Fodor do not discuss the actual intonation pattern used, simply stating that

"the pairs of sentences were recorded on one track of stereophonic recording tape". (p30)

But consider the two following possible intonation patterns for sentences for A and B:

- A) 'in her, hope of, marrying, Anna was' surely impractical
- B) your' hope of, marrying, Anna was' surely impractical

Both readings are plausible, and both clearly signal the appropriate constituent structure. However, acoustically the two common segments are not at all distinct, and in particular the low pre-head on Anna in sentence A is acoustically very similar to the low falling tail on Anna in sentence B. Splicing these two segments of tape might not therefore be expected to produce any great inconvenience to listeners. It is easy to produce plausible intonation patterns of this type for all the sentence pairs used by Garrett, Bever and Fodor.. On the other hand, there is a hint that some of the sentences used involved

rather more disturbing effects, in that the one example where Garrett, Bever and Fodor give a complete breakdown of misplaced clicks shows that a third of the subjects failed to produce a response at all. This is a much higher level of error than would normally be tolerated, and it is surprising to find that no explanation is offered. One probable conclusion is that the splicing produced something so totally unnatural that it severely affected the Subjects' ability to understand the intended sentence. Clearly, then, without a much more detailed account of the intonation patterns and related variables it is difficult to accept these results at face value. Similar criticisms can be made of much of the later click literature, and for this reason, while we subscribe to the view that click displacement may be linked to the size of the perceptual unit, the later developments of the click technique will not concern us here. Rather it is proposed to investigate only the original effect reported by Ladefoged and Broadbent: the difference in accuracy between sentences and digit sequences.

### 3-3. CLICK PLACEMENT AND FOREIGN LANGUAGE LEARNERS.

#### 3-3-1. General Considerations.

If the claim that errors in click placement reflect the size of the unit of processing is correct, then it ought to be possible in principle to use the technique as a way of investigating sentence processing in foreign language learners. In particular we could use the technique in order to test two specific hypotheses. The first of these is a hypothesis that derives from the strong form of the word-list conjecture. If non-native speakers are insensitive to syntactic constraints in the sentences of their foreign language, processing them as if they consisted of strings of words, then these learners should be much more accurate at placing clicks in foreign language material than they are in native language material, and, more particularly, they should be equally accurate at placing clicks in any kind of foreign language material, sentences or word lists alike. The word list conjecture thus predicts the absence of the structural effect normally found with native speakers in the click placement task. However, as we have seen, the word-list conjecture in its strong form is almost certainly too powerful a claim to be correct, the evidence from the previous experiment suggesting that learners do show structural effects, though these are much smaller than those found with native speakers. One way of weakening the word-list conjecture in line with this finding, while still allowing to make reasonably testable predictions, is to allow that learners might be able to make some use of syntactic patterning in sentence processing, but that they generally do not do so. This weaker claim

would lead one to predict that the normal unit of processing for the learner is the individual word, but that in some (unspecified) circumstances, units larger than one word might be assembled and treated as whole units. This "minimal structuring" conjecture makes slightly different predictions from the strong form of the word list conjecture about the performance of non-native speakers on click-placement tasks. If minimal structuring is allowed, then one would expect learners to show a small structure effect, placing clicks more accurately in word lists than in sentences if the material is in a foreign language. However, one would also predict that learners would be more accurate at placing clicks in foreign language sentences than they are at placing clicks in sentences in their native language, since they would generally be using smaller units of processing in the former case. A result along these lines would confirm our rejection of the strong form of the word-list conjecture, and suggest that we ought to be looking for the effects of specific syntactic structures on the performance of learners, rather than pursuing general global structural effects. In practice, as we shall see, there are some difficulties in deciding whether a structural effect in one language is greater than one in another language or not, and this would have lead to some difficulty in interpreting the results of the experiment. In fact, however, this problem was avoided, since the results found are incompatible with either of the two conjectures discussed above, and this outcome is interpreted throwing some doubt on the claim that click displacement phenomena are essentially caused by the size of the unit of processing.

### 3-3-2 Previous work with foreign language learners.

Only one study of click placement by foreign language learners appears to have been carried out previously, an unpublished study by Maidment (Maidment (1973)). This study is a complex one, a design involving a group of native English speakers and three groups of learners of English at different levels of proficiency. Maidment studied the effects of intonation pattern, grammatical type, and the distance of the click from the major constituent boundary on the displacement patterns. . Four measures were taken: a) the number of clicks that were displaced into the constituent boundary: b) the number of clicks correctly located: c) a measure of overall displacement: and d) a measure of clicks displaced towards the boundary, but not into it. Only the first two of these are given any detailed treatment, however.

Maidment's results are difficult to evaluate, as the statistical tests used are incorrect. All the analyses of variance used treat within group variables as though they were between group variables, and since this affects the size of the Mean Square, the resulting F-ratios and significance levels are unreliable. Scores for the individual subjects are not reported, so that it was not possible to rework this data in order to establish the worth of Maidment's claims. Nevertheless, Maidment's results are suggestive, and will be used as a basis for discussion, though the reader should bear in mind that the claims for significant differences may be spurious.

Figure 3-1 below shows Maidment's results for correctly located clicks. (There is a small arithmetical error in column two, which

makes these figures doubly difficult to interpret.)

Table 3-1

Mean percent of clicks correctly located:

GROUP	NS	LP	LI	LB
Click position:				
Before major boundary	28.64	29.47	24.12	23.56
After major boundary	18.39	27.41	16.53	20.01
Combined	23.51	27.13	20.33	21.79

NS = native speakers

LP = learner group (proficiency level)

LI = Learner group (intermediate level)

LB = learner group (beginner level)

Maidment's analysis suggests that there is a significant group effect, and a significant effect due to the position of the click before or after the major boundary. The results are in any case difficult to interpret. The advanced learner group appears to be more accurate at placing clicks than the native speaker group, a finding that is in line with the predictions made earlier, but the two less advanced groups perform at a level slightly worse than the native speakers. Maidment seems to find this result puzzling, in that it implies that learners actually become more accurate in their click placements as they become more fluent, a result that is clearly counter-intuitive. A simpler explanation seems to be that the two weaker groups were simply not able to understand the sentences sufficiently well for the experiment to be a fair test. Maidment actually comments that the Beginner group was not expected to understand the structures.<sup>9</sup>



However, it seems probable that a certain basic minimum level of proficiency is required for the click placement test to be viable at all. At the very least, it is important that the learner should be able to segment the speech signal that he hears into words and relate these to the written forms on the answer sheet. Since Maidemnt's sentences are long, and complex, there is no reason to assume that this task was in the power of the weaker groups.

If then we concentrate on the scores of the native speakers and the Proficiency group of learners, the results seem broadly to indicate that the learners are more accurate at locating clicks than the native speakers are, though the difference is not large, and may not be significant. Rather more interesting, however, is the fact that the native speaker group appear to be considerably more accurate in locating clicks before the major constituent boundary than they are in locating clicks that follow the boundary, while the learner group is not. This suggests that the native speakers are sensitive to at least this syntactic feature, while the learner group is not.

### 3-4. EXPERIMENT .

The experiment reported in this section is a replication of the original Ladefoged and Broadbent experiment, but includes two innovations. Firstly, in addition to the lists of digits and the sentences making up a story, lists of unrelated words were included in the test. Secondly, subjects were tested in two languages and not just in English.

### 3-4-1 Materials.

Two sets of materials were prepared in English and Spanish. Each set consisted of a series of digit sequences, a series of unrelated word sequences, and a series of sentences that made up a short story. For the English set, the digit sequences and the story were taken from Ladefoged and Broadbent (1960). The equivalent Spanish series were closely modelled on these. A complete list of these stimuli will be found in Appendix 3-1.

Each series was recorded on magnetic tape, and a click similar to the clicks used by Ladefoged and Broadbent was recorded over this.

### 3-4-2 Subjects.

Two groups of subjects were tested. A group of native English speakers, who were students of Spanish at Portsmouth Polytechnic or the Polytechnic of North London in their first and second years of study (N=10) and a group of native speaking Spanish students who were following language courses in London at Proficiency level or higher, (N=10).

### 3-4-3 Method.

Subjects were tested individually or in small groups. Each one was given an answer sheet on which the sequences were written, and the sequences were presented one at a time through headphones. Subjects were asked to mark with a slash (/) the place where they thought the click occurred.

### 3-4-4 Scoring.

It was originally intended to mark the answer sheets rather more stringently than in the original paper by Ladefoged and Broadbent. This paper treated each word and the space between words as a slot, and gave one penalty point for each slot the click was misplaced. In Spanish, this measuring is likely to lead to some inaccuracy, as words in Spanish are typically longer than in English, and this might be expected to affect the accuracy of click location. An error of two syllables, for instance, might still leave a click in the same word in a Spanish test, whereas in English stimuli an error of this size would probably place the click in a different word. The obvious measure to use to counteract this problem is the syllable rather than the word, and this measure has been used in some of the click literature. Syllable scoring was used in the pilot stages of this experiment, and produced markedly lower displacement scores for the learners of Spanish than for the native speakers (the mean displacement in syllables on the sentences were 0.59 and 1.26

respectively). In practice, however, it proved extremely difficult to decide which syllable a subject intended a slash to be allocated to. Given haber, for instance, the slash technically belongs to the second syllable -ber, rather than the first ha-, but it seems doubtful whether this degree of accuracy is really justified. Discussion with subjects after the test indicated that such accuracy was not intended on their part. Rather "words and spaces" seems to be an accurate characterization of their perception of the task. Consequently, the difference found in the pilot study was deemed to be spurious, and the analysis used in the main experiment was based on words and spaces.

#### 3-4-5 Results:

Two initial measures were taken: A) the number of clicks accurately reported for each string type, and B) a measure of click displacement similar to that used by Ladefoged and Broadbent but differing in detail. This latter measure is discussed below.

#### Clicks located accurately.

The mean percentage of clicks located accurately by each group on each of the different series is shown in Figure 3-1 and Table 3-2 below. Detailed results for each subject will be found in Appendix 3-2.

Figure 3-1.

Clicks Correctly Located.

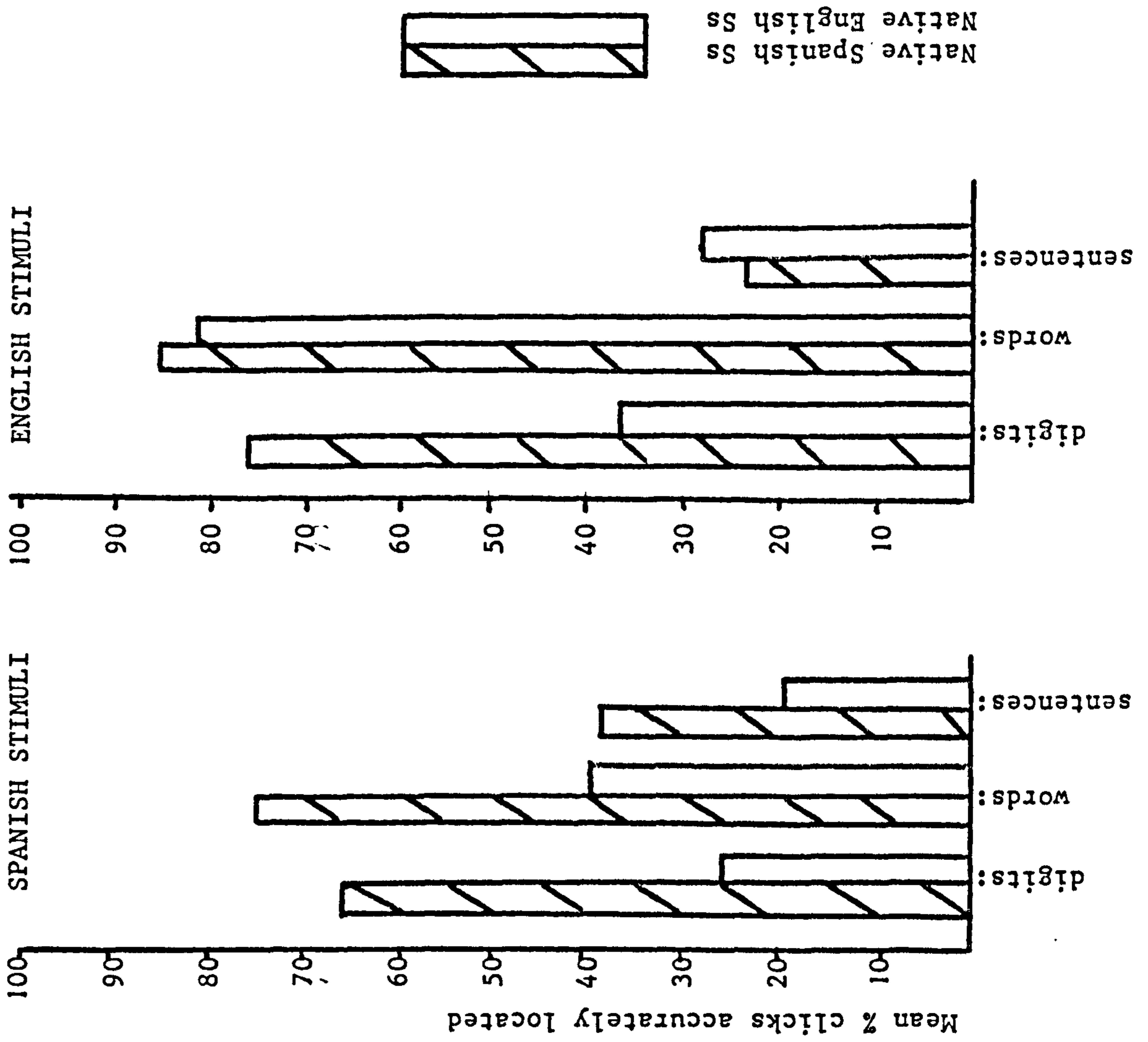


Table 3-2

Mean percent clicks accurately located:

GROUP	Native	Spanish	Native	English
Language of Stimuli	Spanish	English	Spanish	English
<u>digits</u>	65.4	76.0	26.0	37.0
<u>words</u>	75.0	85.0	39.0	82.0
<u>sentences</u>	39.2	24.4	19.3	28.6

The variance of these scores was extremely wide, ranging in some cases from 0-100%, and for this reason a number of non-parametric tests were carried out on the data.

digits: Taking both groups together, there is a tendency for clicks to be more accurately placed in English than in Spanish (Wilcoxon two tailed test:  $T=27$ ,  $N=17$ ,  $P<.02$ ) but this difference is not significant for either group taken on its own. Though the native English speakers are apparently much less accurate on digits than the native Spanish speakers, this difference is significant only for the Spanish stimuli. ( $U=18$ ;  $p<.01$ ) (for the English stimuli  $U=23.5$ ).

words: Here again there is a general tendency for subjects to be more accurate on the English stimuli (Wilcoxon two tail test:  $T=3$ ,  $N=17$ ,  $p<.01$ ), and this time the difference is significant for both groups treated separately (for the English speakers  $T=0$ ,  $N=10$ ;  $p<.01$ : for the Spanish group  $T=2$ ,  $N=7$ :  $p<.05$ ).

sentences: Though the mean scores on the sentence material give the impression that each group scores more accurately in its native

language, these differences are not significant.

sentences compared to words: There is, on the other hand, very clear evidence that subjects consistently score more accurately in word lists than they do in sentences, irrespective of the language of the stimuli. Table 3-3 shows the mean difference scores for the two groups when accuracy on words and sentences are compared; all the differences are significant at the levels indicated.

Table 3-3

Differences in accuracy on word lists and sentences in English and Spanish. These figures are obtained by subtracting each subject's score on sentences from his score on word lists.

	Spanish	English
Native English Speakers:	19.7*	53.9**
Native Spanish Speakers:	25.5**	60.6**

Significance: \* =  $p < .05$ ; \*\* =  $p < .01$ ; (Wilcoxon two tailed tests)

Table 3-3 further shows that the difference we are concerned with is much bigger in English than it is in Spanish. For the English group the difference is significant at the .01 level ( $T=0$ ); for the Spanish group, the difference is significant at the .05 level ( $T=4.5$ ).

#### Amount of Displacement.

Ladefoged and Broadbent's original displacement measure distinguished between clicks moved rightwards and those moved leftwards from their objective locations, in that the former were given a minus sign and



and thus cancelled out movements of equivalent size to the left. This seems to be a rather unsatisfactory way of measuring displacement, as it is possible to obtain a score of zero without placing a single click correctly, if leftward and rightward movement add up to the same total. As long as the number of rightward clicks is small, this possibility can probably be safely ignored. The data recorded here, however, showed a slight tendency for the native English speakers to produce more rightward dislocations in sentences than the native Spanish speakers, a difference which was significant at the 0.05 level ( $U=21.5$ ). For this reason, in the account that follows, all dislocations are counted as positive. A detailed breakdown of all mislocated clicks will be found in Figure 3-5, but a brief summary of this data is provided in Table 3-4.

Table 3-4

Total number of clicks displaced leftwards or rightwards for all sentence stimuli combined:

	left	none	right
English Speakers:	104	41	23
Spanish Speakers:	103	47	8

The basic displacement scores on each string type are shown in Table 3-5 and in Figure 3-2. These scores show a pattern that is closely related to the pattern found with accurately located clicks. This is not surprising, as the two sets of figures are closely dependent on each other. The same non-parametric tests were used to compare these figures as were used in the previous section.

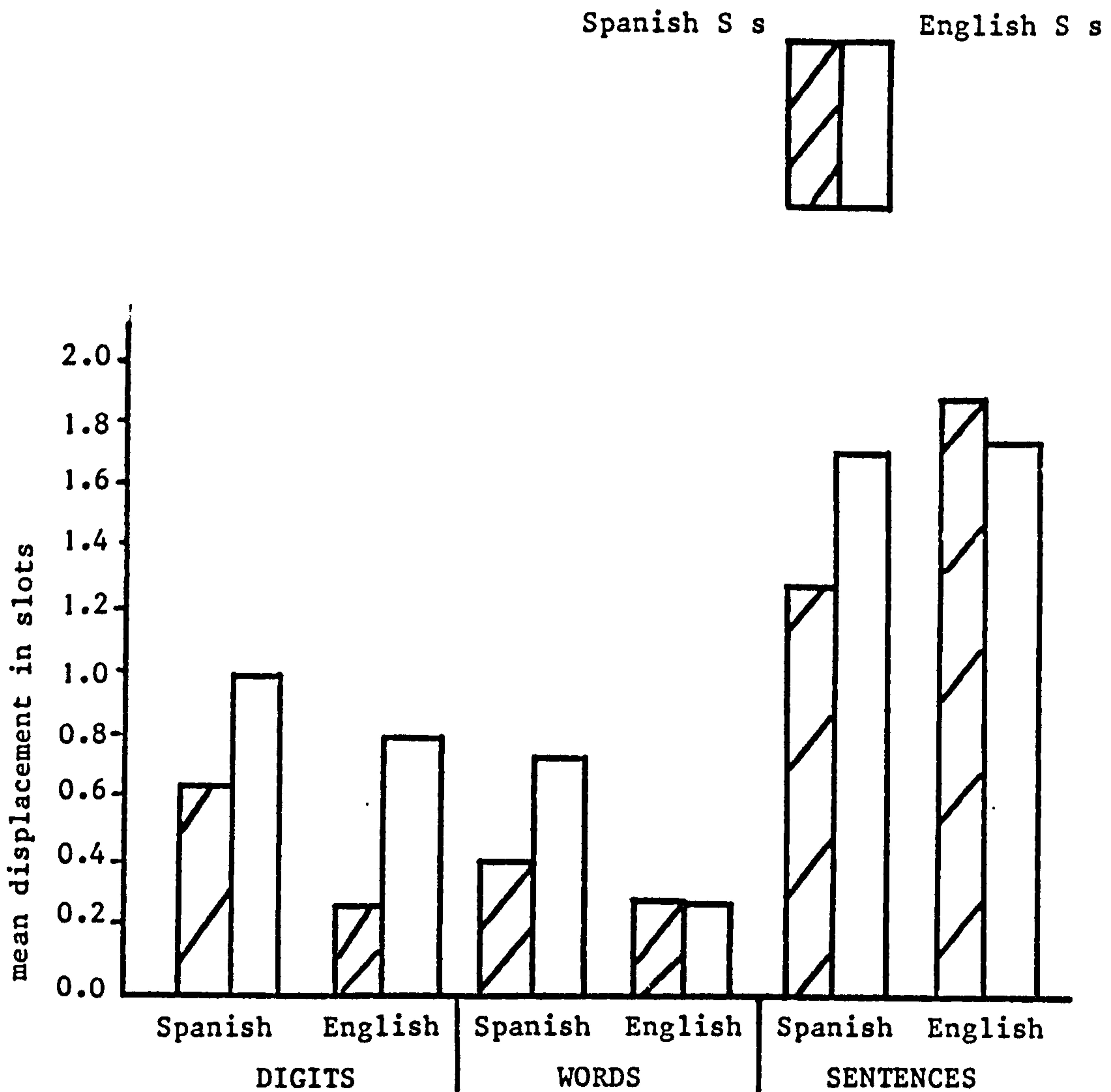
Table 3-5.

Basic Displacement Scores. These scores are derived by calculating the displacement score of each click as described in the text, and dividing the total by the number of stimulus strings.

GROUP	<u>Native Spanish</u>		<u>Native English</u>	
	Spanish	English	Spanish	English
Language of Stimuli				
Digits	0.64	0.30	1.01	0.81
Words	0.42	0.30	0.73	0.29
Sentences	1.29	1.92	1.76	1.79

Figure 3-2.

Basic Displacement Scores.



digits: For the two groups combined, clicks in the Spanish stimuli tend to be misplaced further than in English ( $T=13$ ,  $N=18$ ;  $p<.01$ ), but the difference in scores between the native English and the native Spanish groups are not significant due to the wide variation in the scores within each group.

words: Here again, for the two groups combined, there is a tendency for displacement to be greater in Spanish than in English ( $T=5$ ,  $N=17$ ,  $p<.01$ ) and this difference is also significant for the English speakers considered separately ( $T=0$ ,  $N=9$ ;  $p<.01$ ).

sentences: With these stimuli there was no tendency to score high on the Spanish sentences. Rather the native Spanish group produce significantly wider discrepancies in English than in Spanish ( $T=3$ ,  $N=9$ ;  $p<.02$ ) while for the English group, the two sets of scores are virtually identical.

sentences compared with words: All subjects produced larger displacements on sentences than they did on the corresponding word list, irrespective of language ( $T=0$ ,  $N=10$ ;  $p<.01$  for all cases). The difference appears to be greater in English than in Spanish for both groups, but there is a suggestion that the effect is rather greater for the Spanish group than for the native English speakers. (see Figure 3-3 and Table 3-6).

Table 3-6.

Difference between displacements in word lists and sentences for English and Spanish stimuli. The figures are calculated by subtracting the displacement score on word lists from the displacement score on sentences for each subject and calculating a mean.

Language of Stimuli	SPANISH	ENGLISH
English Ss	0.867	1.717*
Spanish Ss	0.942	1.496*

\* for both groups the second column significantly exceeds the first (T=3, N=10; p<.01)

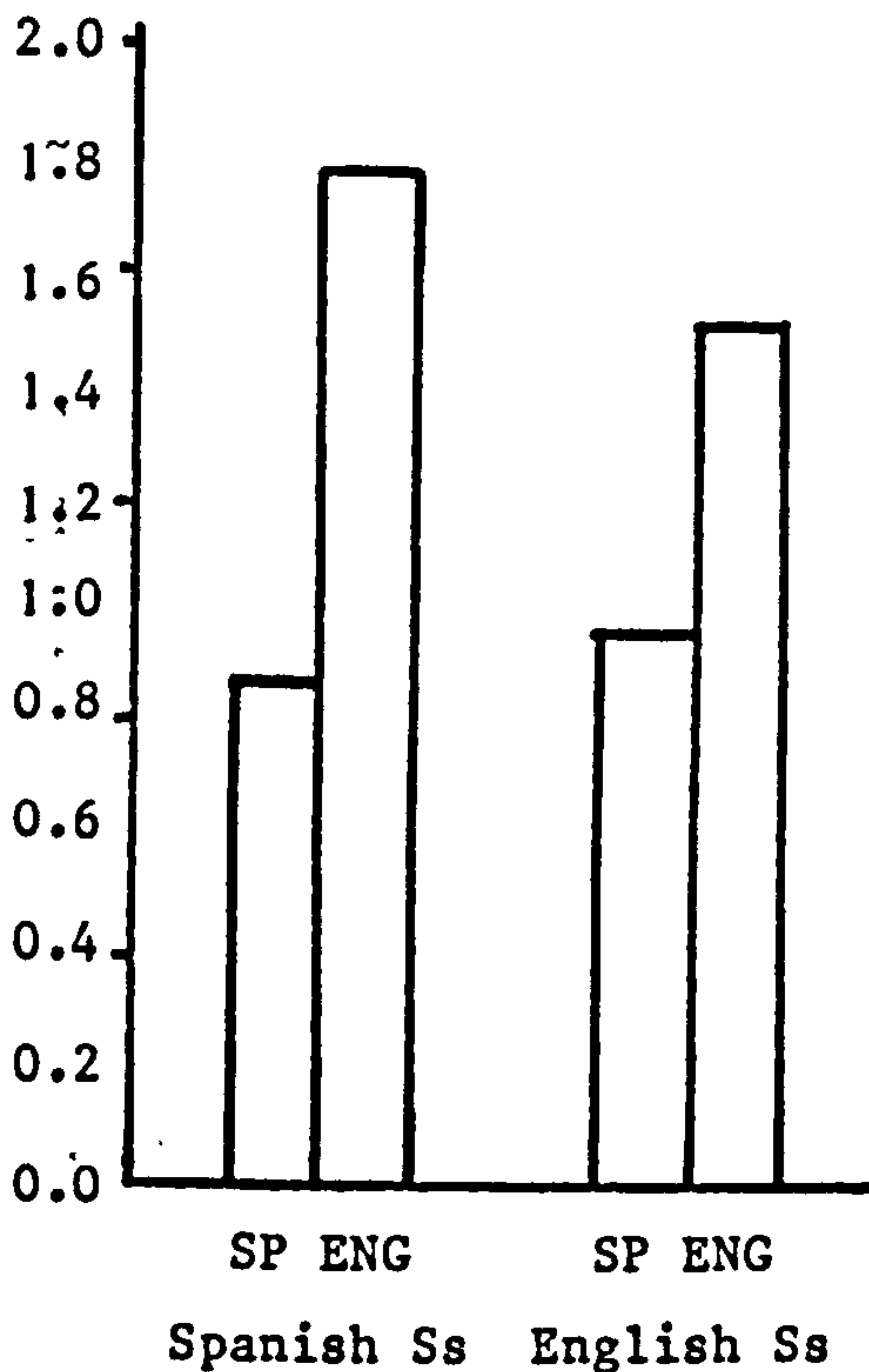
Table 3-7.

Mean Corrected Displacement Scores (see text for explanation).

GROUP	<u>Spanish Speakers</u>		<u>English Speakers</u>	
	Spanish	English	Spanish	English
Language of Stimuli				
Digits	1.8	1.4	1.4	1.2
Sentences	2.1	2.5	2.1	2.4

Figure 3-3.

Difference in Displacement Scores.



A major difficulty with this basic displacement score is that it presents a seriously distorted picture of the movement of displaced clicks if only a few clicks are involved. Thus, in the digit sequence, one click displaced through two positions gives a mean displacement score of 0.2, the relatively large movement of one click being swamped by the other nine clicks that are correctly located. In order to obviate this problem, a second measure was taken: a corrected displacement score. Here the mean displacement is calculated by using as divisor the number of displaced clicks, rather than the number of sequences in the section. The results of this corrected displacement score are shown in Table 3-7, Figure 3-4. The main effect of this score is to reduce the difference between digits and words on the one hand, and sentences on the other, though the mean corrected displacement scores for English words are still misleading due to the large number of subjects who placed all the clicks accurately in this section. The other effect of using the corrected scores is that the heterogeneity of the scores is considerably reduced, and this makes possible the use of more sophisticated statistical tests. An analysis of variance in which the main effects were Group, Language and String Type, was carried out on the corrected displacement scores for digits and sentences in both languages, and the results of this analysis are summarized in Table 3-8. (Scores for the word sequences are still non-normal and were therefore omitted from the analysis.)

Only two significant differences emerged from this analysis: a very highly significant effect due to String type ( $p < .0001$ ) and a small

Figure 3-4.

Mean Corrected Displacement Scores (see text for explanation).

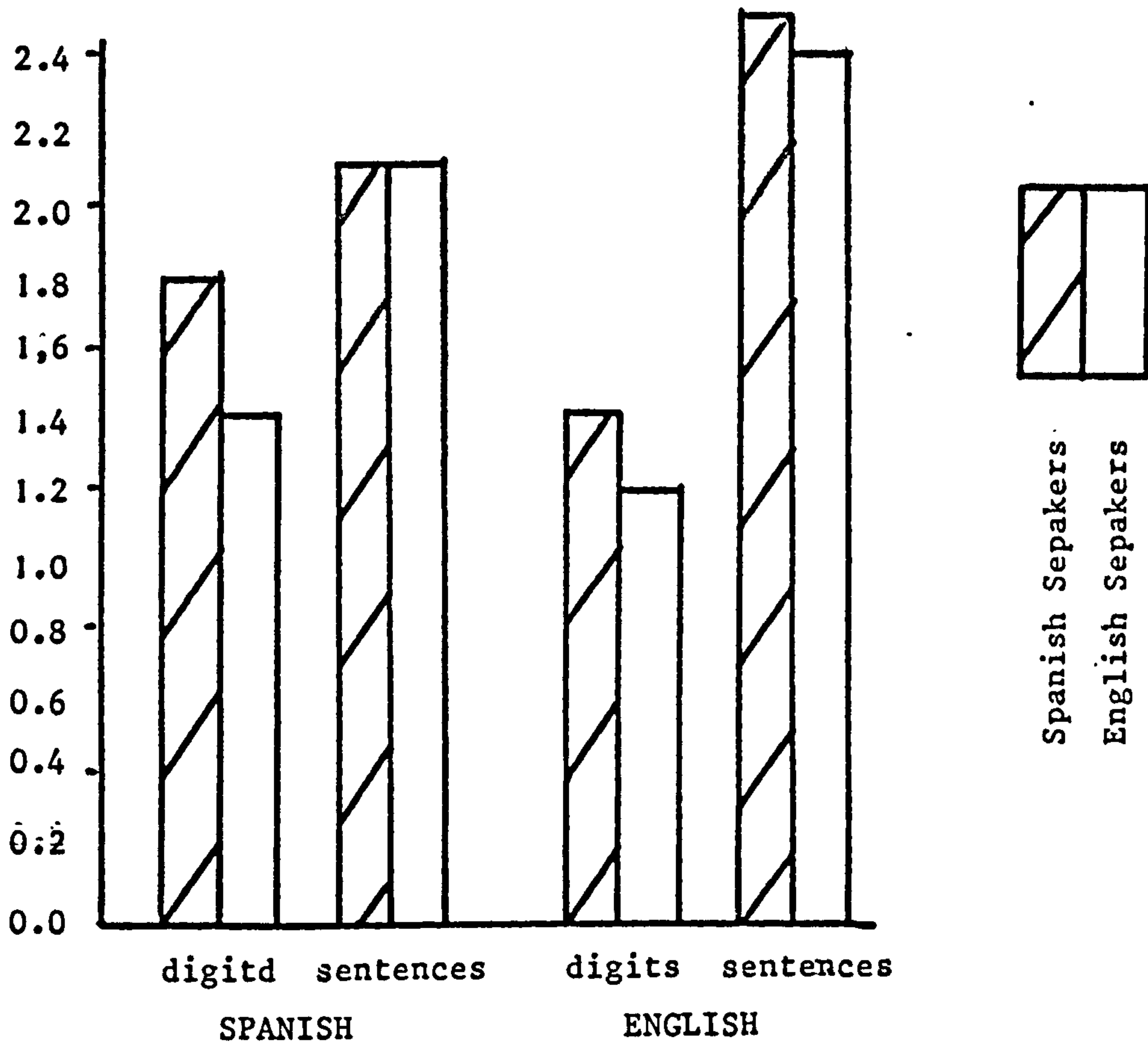


Table 3-8.

Summary of Analysis of Variance.

SOURCE OF VARIANCE	df	SS	MS	F	sig
Subjects	19	9.80			
Group	1	0.72	0.72	1.43	.
Error	18	9.08	0.50		
Language	1	0.02	0.02	0.05	
Language x Group	1	0.01	0.01	0.22	
Error	18	6.45	0.36		
String Type	1	14.11	14.11	43.23	.0000
String Type x Group	1	0.24	0.24	0.74	
Error	18	5.88	0.33		
Language x String Type	1	2.05	2.05	6.03	.024
Lg x StType x Gp	1	0.24	0.24	0.71	
Error	18	6.11	0.34		
Residual Varinace	60	35.1			

interaction between Language and String type ( $p=.024$ ). The String-type effect is clearly due to clicks in sentences being much more displaced than are clicks in digit lists, regardless of language. This finding confirms the analysis of the basic displacement scores. The significant interaction between Language and String-type is rather more difficult to account for, however. The score for all subjects combined on these two variables are shown in Table 3-9. It appears from this table that the interaction is due to a much larger syntactic effect in English than in Spanish.

Table 3-9

Corrected Displacement scores, both groups combined.

LANGUAGE	<u>digits</u>	<u>sentences</u>
Spanish	1.59	2.11
English	1.30	2.46

#### Distribution of misplaced clicks.

The final analysis carried out was an examination of the distribution for each of the individual sentences. Table 3-10 shows the mean displacement score for each sentence. No systematic differences between sentences emerged when the actual distributions of misplaced clicks was studied, though the numbers of clicks involved probably makes analysis of this kind unreliable.

Table 3-11 and Figure 3-5 show the distribution of clicks for all the sentences of each language combined, according to the location of the original click. This table has been simplified in that displacements are shown as percentages of the total number of clicks for each language block. Furthermore, positions further left than

position 3 have been collapsed into a single figure, because of the small number of clicks falling beyond this point. The same procedure has been adopted for all clicks moved rightwards of the objective position. Table 3-11 highlights a number of previously mentioned factors. The greater tendency for native English speakers to postpose clicks and move them rightwards is very clear, as is the tendency for speakers of either language to be more accurate in their native language and to produce a wider range of placements in the foreign language.

Table 3-10

Mean corrected displacement score for each sentence.

SPANISH SENTENCES	Spanish speakers	English speakers	ENGLISH SENTENCES	Spanish speakers	English speakers
1	1.0	2.0	1	2.5	4.4
2	2.25	1.5	2	2.0	1.85
3	2.0	1.7	3	1.6	1.83
4	2.25	2.5	4	2.75	2.22
5	1.3	2.5	5	3.0	2.7
6	2.6	2.2	6	3.7	3.86
7	2.8	2.8	7	1.83	2.0
8	2.5	3.75	8	2.0	2.0
9	1.86	2.5			

Table 3-11

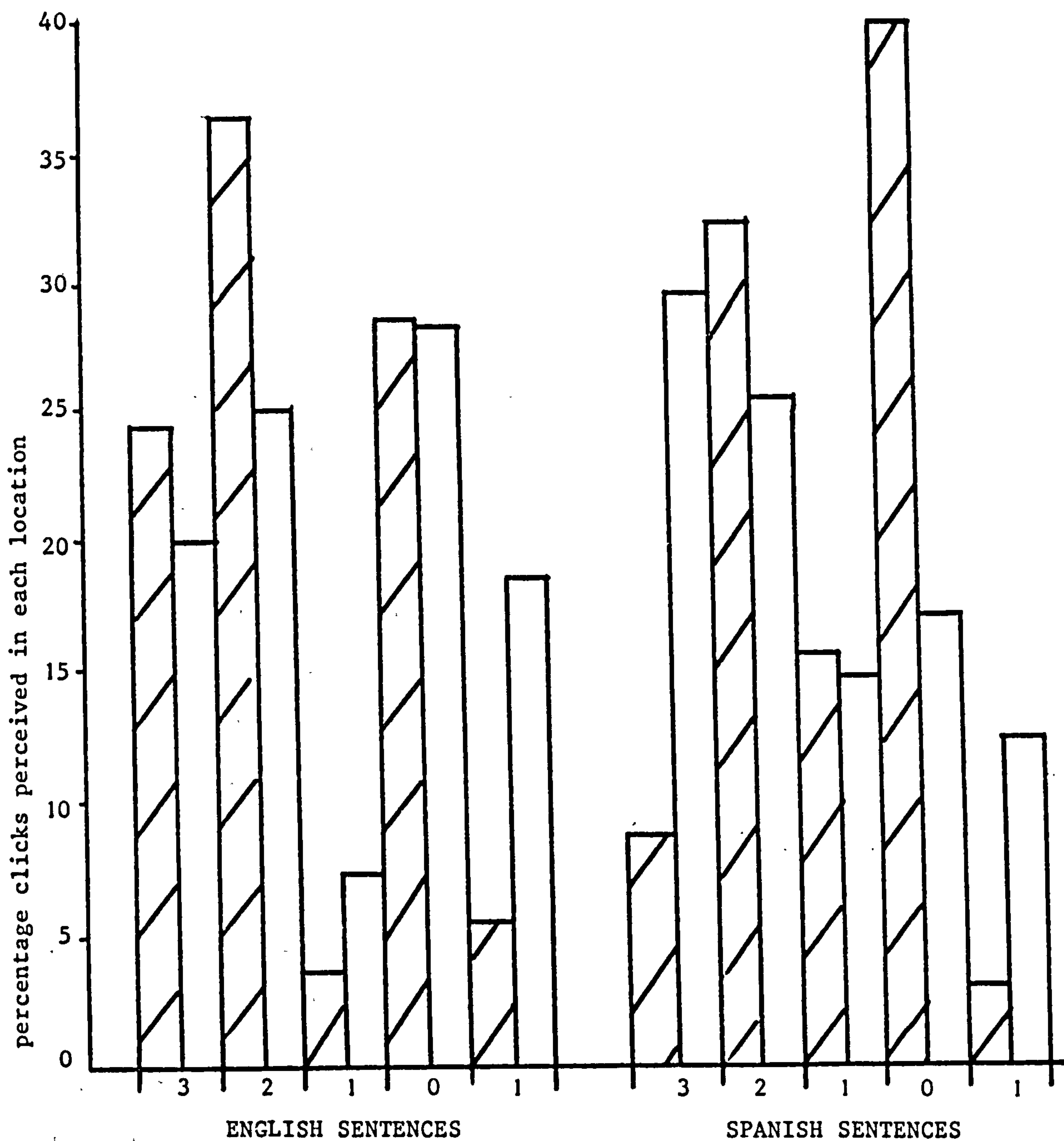
Summary of click displacements.

ENGLISH SENTENCES:						
position:	3+	2	1	0	1+	
English speakers	20.0	25.0	7.5	28.8	18.7	
Spanish speakers	24.7	37.7	3.9	28.6	5.2	
SPANISH SENTENCES:						
position:	3+	2	1	0	1+	
English speakers	29.9	25.3	14.9	17.2	12.6	
Spanish speakers	8.9	32.2	15.5	40.0	3.3	



Figure 3-5.

Distribution of all clicks for all the sentences of each language. Responses of Spanish speaking subjects are shown hatched, English speaking subjects subjects are shown plain.



### 3-5. DISCUSSION.

The main finding of this experiment is that there is apparently a very large structure effect for both groups of subjects, both in English and in Spanish. On the face of it, this finding must count as *prima facie* evidence against both the conjectures that we are considering. The Word List conjecture in its strong form predicted that learners ought to produce minimal displacement in their foreign language, a claim that is clearly falsified. Our second conjecture, the Minimal Structuring Conjecture, lead us to predict that learners should produce less displacement in their foreign language than they do in their native language, but this claim too is falsified by the data, in that there is a clear tendency for learners to produce larger dislocations than native speakers. If we accept that click dislocation errors reflect the size of the basic unit of processing, and that the effect is basically one of perceptual organization, then there is a clear implication that non-native speakers are not handicapped at this level of performance. Any differences between native speakers and learners - for example the differences reported in Chapter 2 - must thus arise as a result of high level processing difficulties; that is to say, that learners have no trouble in accurately perceiving the incoming strings, but they do experience considerable difficulty when they have to store it or perform any other cognitive operation upon it.

However, this rather plausible interpretation of the results actually runs into a number of serious difficulties which suggest

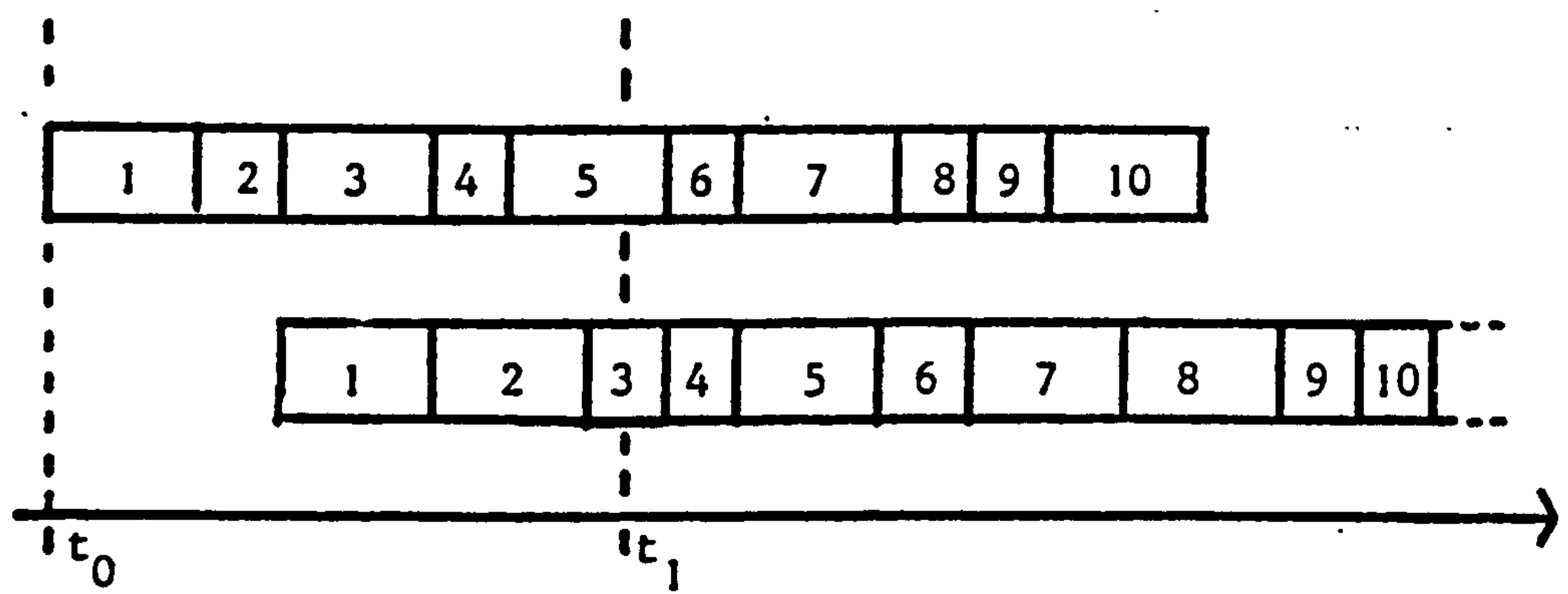
that it may be incorrect. Chief among these is the fact that there are any significant differences at all between performance in one's native language and a second language in the click placement task. If the native speakers and learners both handle incoming strings in essentially the same way, then there is no reason to expect any differences of the kind reported here. The fact that the learners are less accurate than the native speakers, and the fact that learners produce a much wider range of dislocations than the native speakers thus become anomalous findings, for which some explanation is required. Secondly, the direction of these differences is important. If we accept the claim that click placement errors reflect the size of the unit of processing, then these data force us to the conclusion that learners (with larger errors) use larger chunks than native speakers do. This conclusion is clearly counter-intuitive, however, and it is very difficult to think of a theory that would plausibly lead to a conclusion of this sort. The obvious question to ask, therefore, is whether the click displacement effect is the simple syntactic effect it is often claimed to be, or whether there might not be a number of complicated effects involved, each of which interacts with the others.

Consider, for example, the following model. Imagine that when a word is uttered, anyone hearing it requires a finite length of time to decode it and to understand its meaning, so that there is usually a slight delay between the utterance of a word in real time and the listener's grasp of what was said. The size of this discrepancy would vary, depending on such factors as the likelihood of a

given word occurring in the particular context, and this probability would itself be affected by factors such as the frequency of the word, the syntactic environment, and so forth. This process is illustrated in Figure 3-6 below.

Figure 3-6.

(see text for explanation.)



In this diagram, the top line represents a series of spoken words, the length of the boxes giving a rough indication of the objective length of the words. The middle line is intended to give an indication of the complexity of the words heard, the length of the boxes being related to the time necessary to process the word. In short, the top line of the diagram represents real time, while the middle line represents psychological time. Note that the real length of a word is not related to its "psychological length" in any obvious way.

Now let us make a further assumption: non-speech sounds, especially very brief sounds, are perceived instantaneously, and are not

subject to the processing delays that we are hypothesizing for speech. Given such an assumption it is easy to see why clicks will tend to be misplaced leftwards from their objective location. A click occurring at time  $t_1$  in Figure 3-6 is objectively located in the fifth word, but its occurrence coincides with the processing of word number three, and so it is perceived as occurring there, rather than in its true location.

What predictions does this model make about foreign language learners? Firstly, it seems reasonable to assume that learners would have a longer delay in recognizing the words they are listening to, simply because they are foreign words, less frequent, and generally less familiar. This alone would predict that learners should have larger displacements than native speakers. Secondly, if we argue that learners make little use of syntactic structure, paying no heed to redundancies of this kind, again we would expect the recognition latencies of the learners to increase, with the effect of producing even larger mislocations compared with those of the native speaker.

We have now turned the argument advanced in Section 3-3 completely on its head, however. It can only be considered surprising that an experimental technique which has played such an important part in the development of psycholinguistics should be so little understood that it can quite easily lead to quite contradictory predictions.

In addition to this major problem, there are a number of oddities in the data which are puzzling. Chief among these is the

interaction between language and string type discovered in the analysis of the corrected displacement scores. This interaction arises from the fact that the digits in English produce lower displacement scores than digits in Spanish, while sentences in English have rather larger displacement scores than sentences in Spanish. There are no obvious reasons for these discrepancies; indeed one would have predicted that since Spanish digits tend to be disyllables, while English digits are mainly monosyllabic, that Spanish digits would be recorded more accurately than their English counterparts. A one syllable deviation in Spanish is more likely to leave the click within the correct word than a deviation of the same size in English. Similarly, the very large discrepancy between the native speakers' scores in their own language are puzzling. From Figure 3-4, it appears that the native Spanish speakers are as accurate at locating clicks in English as the native English speakers are, whereas the latter group are extremely poor at locating clicks in Spanish. It is not easy to see how this discrepancy between the native speaker groups can be explained. The tendency of the native English speakers to place clicks rightwards of their objective position is also a puzzle, as is the tendency of both groups to place clicks in position 1 for Spanish stimuli, but much less so for English. The current understanding of click displacement does not provide any framework which would allow these puzzles to be evaluated.

CHAPTER 4. THE NEED FOR STRUCTURE SPECIFIC TESTS

In section 3-5 we interpreted the failure of the click placement experiment to produce results that were compatible with either of the conjectures under test as being due to an inadequate understanding of the mechanisms that underlie the click placement phenomena. We suggested that factors other than syntactic ones might be responsible for the large displacement found with non-native speakers - for instance slower word-recognition on the part of non-native speakers might account for the findings, if a model such as that proposed in section 3-5 were adopted. This appeal to word-recognition speeds as an explanatory factor is an interesting one, and will be taken up again in Chapter 7.

In the meantime, however, the questions which we put in section 2.8, and which were our principal motivation for undertaking the click-placement task, remain unanswered. In particular, it is still unclear whether the differences found in Chapter 2 are genuine differences in ability to handle input, or whether they are a product of later, higher level processing deficits in the learners such as an inability to use syntactic structure as an aid in organizing output - what Macnamara (1970) calls "catenation". For the moment questions such as these must remain unanswered. The obvious next step would be to set up another experiment with properties similar to those of the click placement paradigm in that only minimal responses are required of the subjects, and to attempt to resolve the input/output question with another structure sensitive task. Unfortunately, there are no experimental techniques with



anything like the simplicity of the click paradigm which can be used to assess the overall effects of syntactic structure. All the standard techniques for comparing structured and unstructured material make fairly heavy demands on the subjects, particularly in terms of memory load, and would thus leave unresolved the question of whether the differences between native speakers and learners were due to perceptual factors or to response organization. In theory it ought to be possible to devise a new experimental technique that could distinguish between differing claims of this sort, but this is a task which seems beyond the scope of this thesis, whose main concern is to use well-established experimental paradigms as tools to explore the behaviour of a non-standard group of subjects. Consequently, we will leave the input/output question unresolved, while recognizing that this is not a wholly satisfactory state of affairs.

There is in any case, one compelling reason for believing that it might be unprofitable to continue using experimental techniques designed to show that there are gross differences between structured and unstructured sequences of words. We have already shown in both Chapter 2 and Chapter 3 that lists of isolated words in the foreign language are treated differently from word lists that are structurable as sentences or phrases, and to that extent, we have already shown that the Word List Conjecture in the strong form that we have proposed makes incorrect predictions, and is therefore false. Experimental techniques which test only for gross differences between structured and unstructured material would almost certainly confirm our rejection of this conjecture, and might

possibly provide evidence to support the much weaker Minimal Structuring Conjecture, but they seem to be incapable of helping us to refine these notions into some more viable and more interesting claim. Suppose, for example, that we had available an experimental technique that did allow us to measure accurately the size of "chunks" used by learners to process foreign language material. Such a technique might allow us to show that learners had a mean chunk size of 1.2 words, say, a result that would disconfirm the Word-List Conjecture, and as long as the native speaker had rather larger chunk size, confirm the Minimal Structuring Conjecture, but it is not immediately apparent that this gets us very far. The implication would be that learners sometimes use chunks of two words, but not very often, and the interesting question that now arises is what makes it possible for some items to be chunked, but not others. Experiments that concern themselves with global structuring effects seem to be incapable of answering questions of this sort.

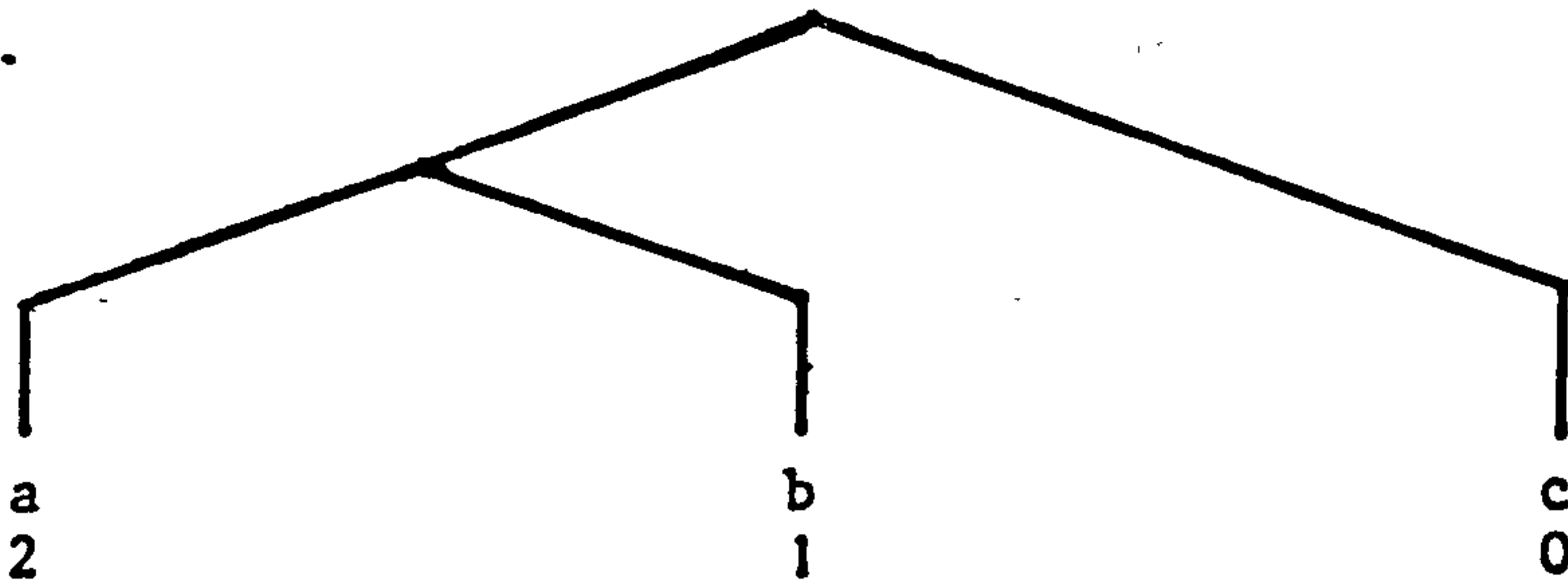
When it is looked at more closely, the notion of a "global structural effect" seems to be a rather unsatisfactory one. The assumption that underlies a construct of this kind is that syntactic structure is a single, unitary property of strings of words, which can be discussed and measured on a macro-scale, and which does not need to be broken down into its smaller constituent parts. This assumption appears to be one which is relatively widespread and pervasive, and it is easy to find in the literature instances of research which look at subjects' ability to use "syntax" or "contextual constraint" or "sequential redundancy" which make no attempt to probe beneath these superficial terms and to enquire more closely

into the effects of particular types of syntagma on performance in experimental tasks. Gerver's (1969) and Salzinger et al's (1964) work on schizophrenics, Craik and Masani (1969) on old people or Carrow and Maudlin (1973) on children are all instances of work of this kind. The possibility that there might be different types of structural relationships and that these might induce measurably different behaviour patterns in subjects is an idea that seems to have been widely overlooked by people working in this area.

To a large extent, the use of tree diagrams - or topologically equivalent forms such as brackettings or triangulated polygons - as schematic representations of syntactic structures, may be responsible for this state of affairs. Though linguists usually distinguish between different types of syntactic relationship by labelling the nodes of their trees as NP, VP, PrepP and so on, it is quite common for psycholinguists to ignore this use of labels. Much of the recent work in experimental psycholinguistics treats any one node in just the same way as it would treat any other node. Several examples of this kind will be found in Chapter 6, where a number of tree diagrams without labels taken from the published work of psycholinguists will be discussed. A more explicit example of this indiscriminate approach to syntactic structure can be found in work based on Yngve's notion of Depth as a measure of psychological complexity. (Yngve (1960), Martin and Roberts (1966) and Greene (1972) especially Chapter 6.) In this type of work, weights are assigned to the terminal nodes of trees without any consideration being given to the type of constituents that the tree is made up of, or even to the type of formative that stands at the terminal nodes; only the formal shape of the tree is taken into account, and differences between

constituent types are discounted. Thus, in the structure shown in Figure 4-1 below, the terminal elements would be ascribed the values shown at the foot of the tree, irrespective of the labelling of the nodes. The only factor of importance is the number of nodes above each formative that have a rightward branch, and their distribution in the tree structure. If the figure were embedded within a larger tree structure, the absolute values ascribed to the terminal elements would be affected, but the pattern of inequalities (a>b>c) would always hold with a configuration of this shape.

Figure 4-1



Thus all plausible realizations of this configuration, such as

(<sub>S</sub>(<sub>NP</sub>the man) (<sub>VP</sub>came) )

or (<sub>Prep</sub>(<sub>VP</sub>sold fish) (<sub>Adv</sub>there) )

or (<sub>S</sub>(<sub>Aux</sub>do not) (<sub>Vb</sub>go) )

or (<sub>NP</sub>(<sub>AdjP</sub>very red) (<sub>N</sub>noses) ) and so forth.

would all be assigned the same values, and treated in the same way, differentiated only by the larger configuration in which the figure might be situated.

Chomsky's own suggestion for a psychological complexity metric based on syntactic structure (which incidentally, does not seem to

have been given any serious consideration by psycholinguists) falls into exactly the same difficulty as Yngve's depth notion. Chomsky and Miller (1963) suggested that the Ratio of Nodes to Terminal Nodes might give some indication of the degree of hierarchical organization of a sentence, this ratio, like Yngve's Depth figure, providing a single figure which could be used to rank sentences on an ordinal scale of complexity. Again, however, the implication is that any configuration of nodes will always have the same value, irrespective of the type of constituents that comprise it.

More recent work in psycholinguistics has begun to question this monolithic approach to syntactic structures - and one major preoccupation of the field at the moment is the psychological effects associated with relatively large constituents, particularly constituents whose boundaries correspond with those of deep structure clauses. It appears that these constituents do have some special properties, functioning very much as if they were perceptual wholes, though given that clauses are fairly obvious candidates for a perceptual unit of this sort, there is surprisingly little evidence available to substantiate even this claim. (See Fodor, Bever and Garrett (1974) and Levelt and Flores d'Arcais (1979) for a review of this work.) This sort of approach to syntactic structure makes an explicit distinction between these larger constituent types and the smaller constituents that comprise them. Implicitly, however, the approach also groups these latter constituent types together into a single undifferentiated set, and fails to distinguish between them in any important way. It is not immediately obvious that this assumption is wholly justified.

There are two main sets of criteria which might be used to provide principled reasons for distinguishing between different types of within-clause constituent. One of these concerns the formal structure of constituents and the rôle they play in sentences. The other is a more complex argument concerned with the relationship between constituent types in two or more different languages. This second argument will be taken up in Chapter 6. For the moment we will concern ourselves only with the first type of dichotomy between constituent types.

In many languages there seems to be a formal distinction that can be made between constituents whose structure is signalled by the occurrence of one of a limited set of function words, and those that are not so signalled. In English, for example, Noun Phrases are often signalled by the occurrence of an article or a demonstrative, Verbal constituents are frequently marked by the occurrence of an auxiliary element, and Prepositional Phrases are invariably introduced by a preposition. All these "marker elements" are short, highly frequent and relatively predictable, and these properties alone should mean that such words are easily recognizable and impose a low cognitive load. If these words act as "flags" for constituents, it would not be surprising to find that the constituent structure of sentences containing a high proportion of such words was much more transparent to subjects than the constituent structure of sentences with a low proportion of marker elements. Indeed, psycholinguists interested in parsing behaviour have suggested that the recognition of marker words of the type we are discussing may be one of the

chief strategies available to subjects as they attempt to identify constituents. (cf. for example Kimball (1973) whose NEW NODES strategy is based explicitly on this principle.) Function words also play a central role in the more formal models of parsing based on Augmented Transition Networks (cf. for instance the work of Kaplan (1972), Bresnan (1978) or Wanner and Maratsos (1978)). The fact that the relationship between such marker elements and the constituents that they introduce is almost always of a single type - viz ((Adjunct) (Head)), an endocentric constituent type is another property that might lead one to expect these constituents to be relatively more simple to handle than more complex exo-centric constituents.

A further reason for distinguishing between these marker relationships and other types of constituent is the fact that in many languages functional markers are not realized as separate words in the surface structure of sentences. Rather they often appear as inflections or affixes on full lexical items, or are expressed through a variety of other devices. One obvious example of this phenomenon is the contrast between English and Latin; in the latter case, inflections on Nouns can often play the same role as prepositions in English - for example in the identification of an Indirect Object. Similarly, Latin frequently inflects its verbs for tense and aspect, where English would typically be able to use a periphrastic form consisting of a verb preceded by one or more Auxiliary forms which are normally counted as separate words. Many examples of phenomena of this type are discussed in Matthews (1974). The usefulness of the distinction even in formal linguistics can be seen from the fact that a similar distinction between full lexical words and

grammatical formatives has at times been proposed for English. Postal (1966) proposed that determiners in English might be profitably regarded not as deep structure formatives, but as features on Deep Structure Noun Phrases, which become realized as separate words relatively late on in the derivation of the sentences. The argument is that from the metatheoretical viewpoint, a feature analysis simplifies greatly the description of syntactic phenomena which are indicated by articles in English. The fact that some languages express definiteness by suffixes, others by proclitics, others by both, and still others by choice of sentence types or ordering can be captured in one metatheory if features are employed.

A similar qualitative distinction between lexical and grammatical formatives seems implicit in much of the discussion of case grammars following Fillmore (1968) - e.g. Stockwell, Schachter and Partee's (1973) suggestion that certain English prepositions can be regarded as case markers.

Where they are realized as separate words, it seems to be frequently the case that markers of this type are chiefly concerned with identifying fairly low level constituent types - and in particular with the isolation of Noun Phrases and Verbs. Much less commonly are they concerned with defining the higher level relationships that exist between the major constituents of a sentence. Prepositions do indeed sometimes serve as markers of oblique case relationships such as benefactive, agentive, locative, and so forth, but determiners in English, and in many other European languages such as French and Spanish typically do not provide information of this sort.



One language for which this claim does not hold, of course, is German, where Determiners not only act as markers for Noun Phrases as they do in English, but also provide information about the relationship of this Noun Phrase to the other major constituents in a sentence through morphological changes in the form of the determiner itself. This might lead one to expect that Determiners in German might play a more important role in sentence processing than they do in English, and there is some evidence to support this claim. The picture is complicated, however, by the fact that in German the information carried by ordering relationships among the major constituents, is also explicitly marked in the morphology of the language, and one set of clues is therefore largely redundant. Whether native speakers of German use morphological clues in preference to word-order clues, providing that all other things are equal, is not clear. A further discussion of some of these points will be found in Mills (1977).

Now it seems plausible to suggest that these low-level structural relationships might be, from a behavioural point of view at least, a rather different sort of thing from structural relationships of a higher order - say, that between a verb and any of the Noun Phrases that depend on it, or even that between a Noun and its dependent Adjective. In fact, when all the differences between full lexical words and grammatical formatives are taken into account, it seems particularly unreasonable to expect that analogies such as:

Noun is to Determiner as Verb Phrase is to Subject Noun Phrase

stand up to close experimental scrutiny.

Surprisingly, however, there is hardly any published work which has looked at differences of this sort in any systematic way.

Almost all the work on the behavioural effects of particular syntactic structures turns out on close investigation to be concerned with instances of higher order structural relations, and not with the effects associated with low-level relationships of the (Det Noun) type.

There are two principal exceptions to this claim. One of these is the work of Levelt (e.g. Levelt (1974)) in which subjects given pairs of words from a sentence are asked to say how strong a connection there is between them. If anything, the results of rating experiments of this sort tend to support the argument advanced here, in that subjects' ratings of what we are calling low level relationships seem to indicate that words in this sort of relationship are perceived as being much more closely related than any other pairs of words. This work is not wholly unequivocal, however, since it does not directly measure the "strength" of relationships between high level constituents - e.g. between the Verb Phrase and the Subject Noun Phrase of a sentence - though indirect measures of this kind are sometimes inferred from a detailed statistical analysis of the ratings for each of the words that comprise the constituents in question. A further problem with this work is that it is not clear that a rating task of this type can be considered to reflect anything that actually goes on in the subject's head when he is handling sentences in more natural circumstances. The absence of any time pressure, in particular, suggests that the rating task may be very remote from ordinary language processing, where efficient storage and real-time constraints of this sort play a crucial role. It is not clear that these covert processes are in any way illuminated by the rating task, and it could be argued that ratings bear

much the same relationship to questions about the psychological structuring of sentences as judgements about grammaticality bear to the psychological processes involved in actually generating sentences.

The second exception to this claim that only high level syntactic relationships have been studied in any detail is a whole series of studies in which the effects of relative pronouns and sentence complementizers on performance have been observed. (cf. for instance, Fodor and Garrett (1967), Hakes and Foss (1970), Slobin (1966) and so forth.) This work is chiefly concerned with the identification of clauses, and the standard experimental technique used is to delete the pronoun or complementizer from a sentence in which it is not obligatory, and then to show that this deletion causes the sentence to be more difficult to process in some way. This type of research is not immediately concerned with structural relationships, and in English, at least, there is nothing comparable to complementizer deletion that takes place within clauses, so that these experiments have no direct bearing on the processing of low level structures. Nevertheless, it could be argued that pronouns and complementizers share many common features with determiners and similar structural words, and the fact that they apparently facilitate handling of strings suggests that it would not be unreasonable to suppose that determiners might play a similar facilitating role within clauses.

If these two sets of exceptions are discounted, it soon becomes apparent that in the great majority of experiments where the effects of syntactic structure have been studied, highly artificial sentence

stimuli were used. These stimuli very often contain no instances of high frequency structural words of the type we are discussing, and where such words do occur in the stimuli, they are often ignored by experimenters, and not treated as part of the experimental data. Some examples of this sort of problem will be discussed further in Chapter 6. It is however, relatively easy to find papers in which example sentences such as:

Snug rings bind chubby fingers (Truscott (1970))

are used without comment. This particular example contains no function words, and therefore no low-level clues to the syntactic structure of the sentence. The only clues available are semantic ones, and higher-level relationships between the constituents. Implausible examples of this kind are by no means rare.

No doubt, the principal reason for this concentration on full lexical formatives is that using sentences that do not contain function words makes it possible for the experimenter to control awkward factors such as word length and word frequency, which are known to effect performance on verbal tasks. It is important to bear in mind, however, that in limiting ourselves in this way, an important distinction may have gone unstudied.

Let us return now to consider the behaviour of non-native speakers of a language. The purpose of this discussion of high-level and low-level structural relationships has been to prepare the ground for the suggestion that non-native speakers of a language might be able to handle the latter type of structure relatively easily, but that they might find it more difficult to integrate

higher-level structures involving major constituents. There are a number of reasons why this is a plausible suggestion. Firstly, as we have seen above, grammatical function words, in Indo-European languages at least, tend to be short and frequent, and this would lead us to expect that these words would impose very little cognitive load on non-native speakers - and certainly a much smaller cognitive load than that imposed by a typical full lexical word. Secondly, grammatical function words are among the first learned and most heavily drilled forms found in foreign language courses, and this intense practice ought to have some measurable effects on behaviour. Another similar reason for ascribing special status to low level structural relationships arises out of the common pedagogical practice of encouraging learners to learn nouns along with a determiner when this serves as a marker of the gender of the noun. It seems very probable that a learner who has acquired "la femme" as a single unit of vocabulary would be able to cope with this structure without much difficulty if it appeared embedded in a larger structure.

The attractiveness of this notion is that it allows us to account for the behaviour of the learner groups in the experiments reported in Chapter 2 without having to abandon the Word-List Conjecture, or rather, it suggests a simple modification to the Word-List Conjecture that goes some way towards describing the behaviour of these groups. This modified Word-List Conjecture consists of the weaker, but more plausible claim, that non-native speakers of a language are able to cope with low-level constituents, but will not show any sign of sensitivity to higher-level structuring between

major constituents. This revised conjecture would lead us to predict that on recall of statistical approximations something very much like the actual results would be found. There ought to be a small improvement on high orders relative to low ones, and this improvement would be mainly due to the presence of function words which act as pointers to low-level constituents. The conjecture also allows to explain why the learner groups tend to do better on 2nd and 4th order approximations than they do on higher orders. Typically, higher orders such as 5th and 6th orders of approximation are characterized by quite long stretches of normal prose, made up of whole well-formed clauses, and exemplifying higher level structural relationships. By contrast, orders of approximation in the 2nd to 4th range typically contain Noun Phrases or Prepositional Phrases, but it is rare to find a whole well-formed clause. Consequently, if learners can handle the low-level clauses easily, but not the higher ones, one would expect to find a slight decline in performance on the higher orders relative to the middle ones. Some further support for the revised conjecture comes from the sequence scores reported in Section 2-7-1. Here we saw that there was very little evidence for recall of sequences of more than two or three words. Recall of much longer sequences would be expected if higher level structures were being handled with any degree of fluency.

These findings are suggestive, and lead some credence to the revised Word-List Conjecture. However, it seems that the immediate recall paradigm is really too coarse a measure to enable us to test the revised conjecture with any degree of confidence, and in order to do this we must move into another type of experiment that uses a

rather more sensitive technique. An experiment that investigates recognition thresholds for tachistoscopically presented material will be reported in the next chapter. This experiment provides some further support for the revised Word-List Conjecture.

CHAPTER 5. RECOGNITION THRESHOLDS FOR PHRASES

5-1 Introduction

5-2 Previous Studies of Word Recognition in a Second Language

5-3 Pilot Study

5-4 Main Experiment

5-5 Discussion



## 5-1. INTRODUCTION AND BACKGROUND.

The work reported in this chapter is concerned with Tachistoscopic Recognition Thresholds for word sequences of different types in two languages. The work follows on from the discussion of the previous chapter, in that it is concerned with specific types of syntactic structure, rather than with syntactic structure as a global, undifferentiated property of strings of words. Recognition of tachistoscopically presented words was chosen as the experimental task because it makes very simple demands on subjects, and to that extent it differs from the work described in Chapter 2, and avoids some of the problems of interpretation that arose out of that work. Word recognition puts a premium on rapid processing of input material, but does not make very heavy demands in terms of output. Short Term Memory capacity, speaking abilities, and so on do not play any crucial part in the test. To this extent, the recognition threshold task has rather more in common with the click placement task discussed in Chapter 3.

The experimental literature on word recognition by Native Speakers of English is very large indeed, and for this reason it is not proposed to review it here. The authoritative summary of this work is Gibson and Levin (1975) - though cf. Coltheart (1977) for a fairly critical review of Gibson and Levin's theoretical framework. A detailed account of some specific factors affecting thresholds for individual words is to be found in Morton (1978), which is particularly concerned with frequency.

Surprisingly, the number of studies which have looked at the perception of sequences of words rather than individual words is remarkably small. Pioneering work in this field was carried out by Cattell in 1885. He reported that

"numbers, letters, words and sentences were exposed to view for 0.1 seconds, and it was determined how much consciousness can attend to at one time . . . . . On average, consciousness can grasp four numbers, three to four letters, two words, or a sentence composed of four words . . . . . Twice as many words can be grasped where they make a sentence as when they have no connection." (p311)

Despite the fact that Cattell's work used fairly unsophisticated equipment consisting mainly of a home-made magic lantern, these findings do not appear to have been extended or developed in any serious way, and Cattell's work is still cited as standard, both in the textbooks of the inter-war period (e.g. Woodworth 1938, p738) and in more recent work such as that of Gibson and Levin (1975, p189). The fact that Cattell's claim that consciousness can grasp four numbers, or three to four letters is now known to be a gross oversimplification and that far more information than this is available from even very brief tachistoscopic displays (cf. Sperling (1960) and Neisser (1967)) does not appear to have provoked any similar re-evaluation of the claims made for apperception of connected material.

## 5-2. PREVIOUS STUDIES OF WORD RECOGNITION IN A SECOND LANGUAGE.

Only a handful of studies investigating the perception of words in a foreign language have been carried out. Again, the earliest work in this field is that of Cattell (1885) who found that recognition thresholds for single words was higher in a foreign language than in the subjects' native language. This finding is difficult to interpret, however, since the claim is based on a study of only two subjects, one of whom was Cattell himself. A further reason for treating the claim with some caution is that the language studied was German, and in the 1880's German would commonly have been written with a Gothic script rather than in one of the more common varieties of the Roman Alphabet. It is well known that Gothic is not an easy script to read, and so it is not clear whether the differences reported are genuinely due to material being in a foreign language, or to the unfamiliarity of the script they were written in.

The question of unfamiliar scripts is one that frequently intrudes into threshold experiments where second languages have been studied. Albert and Obler's (1979) review of research into word recognition by bilinguals refers almost exclusively to tests where not only the written forms of letters, but even the basic principles on which the writing is organized differ quite radically from the writing conventions associated with Western European writing systems. Albert and Obler cite Dalrymple-Alford's (1967) paper comparing English and Arabic; Mishkin and Forgay (1952), Orbach (1953 and 1967) and Barton,

Goodglass and Shai (1965) who compared the perception of Hebrew and Yiddish words with English; and Kershner and Jeng (1972) who found broadly comparable patterns of perceptual behaviour in both English and Chinese.

All this literature is primarily concerned not with establishing whether there is a basic difference in thresholds for material in a second language, but rather with the idea that there might be some differences in the way second language material is processed by the brain. In particular, the authors cited by Albert and Obler are all attempting to show that the right visual field effect normally associated with printed words - i.e. words presented in the right visual field are usually processed more effectively than words presented in the left visual field, if the subjects' dominant hemisphere is his left one - might be less marked for second language material. This basic question undoubtedly accounts for the predominance of studies that have looked at Semitic scripts, since the typical right to left letter order of these scripts might be expected to produce a left visual field advantage if anything is capable of doing so.<sup>10</sup> The rather different preoccupation of these studies means that much of this work is tangential to our own interests, however. Also marginal, but for rather different reasons, is the work of Lukatela et al (1978) in which the perception of Russian words written in the Cyrillic Alphabet was studied. These authors are particularly interested in the effects of unfamiliar letter forms in a sequence which otherwise consists of familiar letters, and the effects of familiar letters with radically different values ascribed to them.

One of the few papers to discuss recognition thresholds for single words in a foreign language that uses a relatively uncomplicated Roman script - for Lukatela et al.'s problem with letters being assigned different values crops up in any second language to some extent - is that of Lambert, Havelka and Gardner (1959). Lambert, Havelka and Gardner's experiment compared the recognition thresholds of three groups of subjects for twenty words in English or French. The groups consisted of bilinguals who were either balanced or dominant in one of the two languages, a characteristic which is defined in terms of reaction times on a simple motor task (cf. Lambert (1955) for details.) It is with this latter variable that Lambert, Havelka and Gardner are mainly concerned, and as a result, their account of the experiment is very brief; details are sketchy - for example we are told only that "the words varied in content between languages, but were controlled for length and frequency of occurrence"; the frequency range and the length are not provided however. The results reported are cryptic in the extreme:

"It was our prediction that the closer bilinguals approach balance in the test of speed of responding, the more likely they are to have similar recognitive thresholds for words in their two languages. Average thresholds for each language were determined and difference scores between languages (French exposure time minus English exposure time) computed for all subjects. The product moment correlation between degree of bilingualism and comparative recognitive thresholds is 0.46, significant at beyond the 1% level. It appears then, that the degree of bilingualism is reflected in the perceptual processes." (p65-66)

One can probably infer from this that the group which was dominant in English had higher thresholds for French words than for English

words, and vice versa for the dominant French group, but this is not wholly certain. The correlation figure is logically independent of the actual level of the thresholds, and is sufficiently low to allow for the possibility that there may be no significant difference between the groups on the recognition thresholds themselves.<sup>11</sup>

### 5-2-1 Macnamara.

The main source of data concerning the ability of foreign language learners to apprehend sequences of words in their foreign language comes from a series of experiments by Macnamara and his colleagues. There are no wholly satisfactory accounts of these experiments, in that the details of the tests are never fully discussed, the full results do not appear to have been published, and there is no proper statistical treatment of the data. However, as Macnamara's work is widely known, and has often been cited uncritically as lending support to the claim that non-native speakers make no use of syntactic redundancies in their perception of foreign language material (e.g. by Albert and Obler (1978) p207) this work is considered here in some detail. The best accounts of the experiments are to be found in Macnamara (1967) and Macnamara (1970) and the discussion below is based on these two sources.

Macnamara's first experiment was concerned with the performance of a group of English-dominant adolescent girls (N=24) in a series of simple tasks in English and French. Four sets of tests were carried out. These are listed in Table 5-1 below, and Table 5-2 shows the results on each task.

Macnamara writes: "the mean difference between perceptual thresholds for English and French words is not significant. Perceptual thresholds for English and French sentences differ significantly. However, if they are corrected for individual differences in thresholds for words, the difference for sentences falls short of significance. Thus, it seems reasonable to conclude that

Table 5-1

Tasks examined by Macnamara (1967) and (1970).  
Each task was carried out in both English and French.

- 1: Tachistoscopic recognition thresholds for words.  
Tachistoscopic recognition thresholds for sentences.
- 2: Reaction times for matching words to pictures.  
Reaction times for assigning truth values to sentences.
- 3: Silent reading of scrambled text.  
Silent reading of canonical text.
- 4: Reading aloud of scrambled text.  
Reading aloud of canonical text.

(All the stimuli used the same basic materials, in that the sentences are composed of the words used in Tasks 1 and 2, and the texts used in tasks 3 and 4 are composed by concatenating these sentences. Scrambled text is constructed by disordering the words in each sentence so that the syntactic structure of the sentence is destroyed.)

Table 5-2.

Time taken to complete the tasks listed in Table 5-1.  
(from Macnamara (1967)).

TASK	ENGLISH	FRENCH
<u>Recognition Thresholds</u>		
words	69 m.secs	75 m.secs
sentences	236 m.secs	270 m.secs
<u>Naming and Assigning Truth Values</u>		
words	1089 m.secs	1230 m.secs
sentences	1448 m.secs	1704 m.secs
<u>Silent Reading</u>		
scrambled texts	12.01 secs	12.92 secs
canonical texts	7.37 secs	10.00 secs
<u>Reading Aloud</u>		
scrambled texts	17.70 secs	20.54 secs
canonical texts	9.64 secs	14.80 secs



the recognition of sentences brought into play no skill not involved in the recognition of words." (p128)

He goes on to isolate four components in tasks three and four:

- a) perception of individual words
- b) pronunciation of individual words
- c) use of transitional probabilities
- d) catenation (i.e. linking the words together in speech).

Macnamara argues that the various reading times are each affected by these separate components as follows:

silent reading of scrambled passages .....: (a)  
reading scrambled passages aloud .....: (a+b)  
silent reading of canonical passages .....: (a-c)  
reading canonical passages aloud .....: (a+b-c-d)

Since factors (c) and (d) reduce reading time, they are counted as minus qualities.

Some simple arithmetical manipulations make it possible to estimate the size of each of the separate components, and the contribution that each one makes to the overall performance time. Macnamara then calculates the difference between the contribution of each component to performance in English and French. "Significant differences" - the size of the difference, and the level of significance are not recorded - were found only in relation to components (b) and (c), and Macnamara interprets this latter finding as indicating that subjects "made less use of the transition probabilities in French." Macnamara (1970, p113).

Some caution is needed with this claim, however. Macnamara's method of isolating the components that he is interested in seems to be rather indirect and unreliable. The crucial figure in the calculations -

crucial in the sense that it appears as a factor in all the other expressions - is the one for silent reading of scrambled passages. The very idea of asking people to "read" a passage of scrambled sentences seems to be of dubious validity, however. The task is only marginally related to normal reading, and it seems unreasonable to assume that the same basic processes are involved in these tasks. It is impossible, for instance, to read a scrambled passage in order to extract its meaning, as a scrambled text of this sort is essentially meaningless. There are, furthermore, considerable difficulties in making accurate measurements of silent reading speed. Macnamara's method consisted in asking his subjects to point with their finger to the word they were reading, a technique which does not seem designed to produce highly accurate reading times. The measure is subject to sizable errors on the part of both the subject and the experimenter, and in many cases preempts the subject's decisions, by forcing him into a linear word by word approach which may be inappropriate to the material being read.

A further set of problems with these data comes from the type of material Macnamara uses. The sentences are all simple sentences of a single syntactic type, of the form

"A \_\_\_\_\_ has a \_\_\_\_\_."

and their French equivalents, where the blanks are replaced by one of a set of eight nouns. These eight nouns were also used as the stimuli for the matching task and the word recognition task. The reading passages were constructed by deleting the full stops and linking the sentences together with and or et as appropriate. Scrambled passages were constructed in the same way using the scrambled sentences linked

together by and or et.

Macnamara gives as examples of the sentences used "A hen has a wing" and its French equivalent "Une poule possède une aile." (sic.) The first thing to note here is the odd use of posséder in French. It would be more normal to use avoir, but even then the sentences still sound odd in either language. A more serious problem is the fact that possède is more than twice as long as has, and considerably less frequent. Both these discrepancies should have been taken into account in an experiment that is concerned with small differences in reaction times, as both frequency and length are known to affect recognition thresholds. A further serious problem is that only one type of syntactic structure is used in all the sentences. This means that the syntax of the sentences is entirely predictable and the importance of transitional probabilities reduced to a minimum.

Macnamara appears to be aware of some of these shortcomings, and in a further experiment reported in his 1970 paper, he used a rather wider range of syntactic structures, though the total vocabulary remains unchanged. He writes:

"In order to verify our explanation of the absence of a significant difference associated with syntax in the interpretation of sentences, a second experiment was carried out. This time syntax was systematically varied so that the subjects would have to pay attention to it. The new sentences were of four types: active affirmative, active negative, passive affirmative and passive negative. In composing the sentences, however, we did some violence to both English and French syntax. For example, one set might read:

a hen possesses a wing  
a hen does not possess a wing  
a wing is possessed by a hen  
a wing is not possessed by a hen

The corresponding set of French sentences would be:

une poule possède une aile (sic)  
une poule ne possède une aile (sic)  
une aile est possédée par une poule (sic)  
Une aile n'est pas possédée par une poule (sic)

We were particularly worried by the violence done to French syntax. Nevertheless, as the results indicate, the sentences served our purpose well enough." (p113)

It is difficult to know where to begin to criticise these materials. All the French sentences contain gross grammatical errors or spelling mistakes. The use of posséder with inalienable properties such as wings is bizarre in French, and equally disturbing in English. Furthermore, the use of the passive form and the negative-passive is highly restricted in French, and certainly not permissible here, so that the French versions of the English stimuli are not really equivalent at all. In short these French stimuli and the comments that follow them show an extraordinary lack of sensitivity to the French language. Macnamara does not provide detailed results for this second experiment. He claims, however, that an essentially similar pattern of results emerges as was found in the first experiment:

"the increase in time from the words and pictures task to the sentence task was significantly greater in the weaker language, and significant differences between languages were found ..... in ability to anticipate sequences of words in continuous prose."  
(p114)

### 5-2-2 Morris (1978)

A study much more reliable than that of Macnamara is Morris (1978). In this experiment, a group of native English speakers and a group of foreign learners of English were asked to read four-word sequences of English words in a single tachistoscopic exposure of 100 m.secs. The word sequences were of two types: simple four word English sentences, or scrambled versions of these sentences which could not be construed as sentences. Each subject read eight scrambled forms and eight canonical forms, and the scrambled versions of each sentence were read by half of each group, while the other half saw the canonical version. Morris' results are shown in Table 5-3 below.

Table 5-3.

Mean number of words correctly reported by native speakers and learners of English at an exposure of 100 m.secs.

String type:	Canonical	Scrambled
Native Speakers	26.76	17.58
Non-native speakers	7.5	4.75

An analysis of variance indicated that both the main effects, Group and String Type, produced very significant differences ( $p < .001$ ) There is also a small interaction between these two variables, which seems to be due to the learner group showing a much smaller improvement on canonical forms. Small though it is, however, this difference is still significant at the 5% level, and this finding seems to indicate that the learner group do apprehend more words in the canonical sequences than they do in the scrambled ones. The importance of

Morris' evidence lies in the fact that her subjects were not advanced learners of English. Rather, they were chosen on the grounds that they had all failed to score more than 8% on a simple Cloze Test. Morris' results suggest that even learners with a very elementary knowledge of their foreign language perform better on canonical sequences than they do on scrambled ones, and this evidence seems to conflict with Macnamara's claims.

There is, however, one aspect of Morris' data which is puzzling and makes her interpretation of the results doubtful. Though the learners' scores on the canonical sequences are nearly 60% higher than on the scrambled sequences, this figure still represents an average apprehension span of less than one word. It is difficult to see how a structural effect, which by definition requires two words at least to be effective, could come into operation under these circumstances. To draw a parallel from the field of letter recognition, Morris' results are as if a group of subjects produced higher scores on words than on jumbled letter sequences, but still reported only one per exposure. It would be very difficult to argue for the existence of a genuine word effect given such a result, and one would be inclined to look for other possible sources of improved scores, such as letter frequency. An explanation along these lines may possibly account for Morris' results.<sup>12</sup> Morris' 1978 paper does not list the sentences used in her experiment, but a complete list is to be found in an earlier paper (Morris (1977) p50). Morris states that all the words chosen were taken from the Thorndike-Lorge list of words with a frequency of over 100 per million. It appears from the list

of stimuli, however, that the canonical sentences almost all begin with short function words such as this, that, the, her, etc. Now, though these words technically fall within the same frequency class as very frequent nouns, adjectives or verbs, they in fact belong to the class of determiners, one of the groups of grammatical function words that we discussed in the previous chapter. There we suggested that very high frequency items of this type ought to be more easy to handle than full lexical items, simply because they are typically shorter than other words, and because they are carefully drilled and practiced by second language learners. It appears then, that Morris' frequency control, though formally sufficient, may have been ineffective since the Thorndike-Lorge count does not distinguish between words falling in the AA class. This means that the results of this study may not represent a genuine structural effect after all, but rather an artefact due primarily to uncontrolled word frequency effects. Even if this argument is rejected, and the validity of Morris' frequency effect accepted, then it is still the case that the whole of the learners' improvement on structured sequences could be ascribed to the low-level effects we discussed in the previous chapter, since the first two items of her structured strings are always instances of a (Determiner Noun) syntagma.

### 5-2-3 Summary.

Neither Morris' work, nor the work of Macnamara are incompatible with the revised Word-List Conjecture which was advanced in Chapter 4. Macnamara's claim that learners do not make any use of transitional probabilities between words when they are performing in their foreign language is actually rather stronger than the Revised Word List Conjecture, and lies much closer to the original Word List Conjecture discussed in Chapters 1 and 2. However, as we have seen, Macnamara's findings are highly suspect, and it would be unwise to take them at face value, or to advance them as evidence in support of any claim. Morris' data at first view contradicts the Revised Word-List Conjecture, but her apparent structural effect seems to be either an artefact due to word frequency effects, or a genuine structural effect which is limited to low-level structural relationships of the Det-N type. If either of these explanations are accepted, then Morris' data ceases to be an embarrassment. Indeed, given the nature of her stimulus strings, the Revised Word List Conjecture would predict an improved score on structured strings over scrambled ones that is very much in line with the actual findings.

Neither Morris nor Macnamara provides a really clear test of the revised Word-List Conjecture, however, and so fresh experimental evidence was sought in an attempt to provide a clear refutation of the conjecture. This work is reported in the following sections.



### 5-3. PILOT STUDY.

The aim of this pilot study was to test whether reliable differences could be found between different types of visually presented material - to confirm, in effect, Cattell's claim that structured material is more easily perceived than unstructured material in one's native language, and to search for signs of different behaviour in a foreign language.

### 5-3-1 Materials.

Two sets of material were prepared, one set in English, the other in Spanish. Each set consisted of sixteen word pairs, four Noun-Noun pairs, four Adjective-Adjective pairs, four Verb-Noun pairs, and four Adjective-Noun pairs. (In the Spanish stimuli this last set was replaced by four Noun-Adjective pairs, because the normal Spanish word order in Noun Phrases differs from that of English.) The main difference between the pair-types is that the Noun-Noun and Adjective-Adjective sequences are not construable as constituents, whereas the remaining pairs are. The Verb-Noun sequences could all be interpreted as Verb Phrases with the noun acting as the Object of the Verb; the Adjective-Noun/Noun-Adjective sequences could all be interpreted as Noun Phrases. All the words chosen were high frequency items, five letters in length. The full list of stimuli will be found in Appendix 5-A.

The word pairs chosen for study do not contain any examples of Det-Noun sequences. This may seem odd in the light of the earlier discussion in which we criticised other authors for ignoring structures of this type, and some explanation for our choosing to follow the same course of action is clearly necessary here.

The Revised Word-List Conjecture makes two logically independent claims:

- a) that higher-level structures are treated as though they were unrelated word-lists, and
- b) that low-level structural relationships involving highly

frequent, short grammatical function words may be structured in the normal fashion.

There are three principal ways in which one might set out to test these claims. One is to compare the performance of learners on unstructured word lists and low-level structures of the Det-Noun type. The conjecture predicts that these latter strings should be handled more easily than the former, and that the difference would be clearly related to a similar difference in the performance of native speakers. The second approach would consist of a test to compare the performance of learners on low-level and high-level structures - say, a comparison of (Det-Noun) structures with (Noun-Verb) structures. Here the conjecture predicts that the former structure type would be handled more easily by learners than the latter, and that this behaviour should be markedly different from that of native speakers who might be expected to handle both types of structure equally easily. The third approach is to concentrate on claim (a), ignoring the low-level structures, and focussing instead on a comparison between high-level structures and unstructured word-lists. It is this last course of action that is adopted here. The main reason for this is a practical one. In word-recognition experiments of the type used in this chapter, controls on length are obviously crucial, but it is simply not possible to find a sufficiently large number of determiners of any one length either in English or Spanish to produce a consistent set of stimuli within any given length constraints. Using the same function words several times over is not an acceptable way of solving this problem, since repeated exposure of the same word results in a marked

lowering of its normal threshold. Similar difficulties arise if one attempts to control for frequency, as there are hardly any full lexical items with a frequency of occurrence that is comparable with the frequency of occurrence of grammatical function words, and this again makes valid comparisons difficult. Practical considerations such as these mean that the Tachistoscopic Recognition Threshold Technique is not suited to testing claims based on the second part of the Revised Word-List Conjecture, though it can be used easily to test claims that are based on the first part. For the moment, then, the claims about high-frequency function words must be considered only as interesting ideas which lack any solid experimental support.<sup>13</sup>

In one sense, however, the fact that technical considerations such as these pre-empt our options may not matter very much. The criticism of those who ignored the distinction between high and low-level structure types was made in the context of authors who do this but fail to acknowledge that they have done so, and continue to talk about "syntactic structure" without explicitly detailing the structures which they have investigated, and those which have been systematically ignored. This mistake has been avoided here, since we state explicitly which structures are being tested, and no structural claims of a global kind are being advanced.

A number of other technical problems arose in the preparatory phases of this study. The first of these is that Spanish uses an alphabet which is slightly different from the English one of twenty-six letters, in that it includes two digraphs: ch and ll. These digraphs

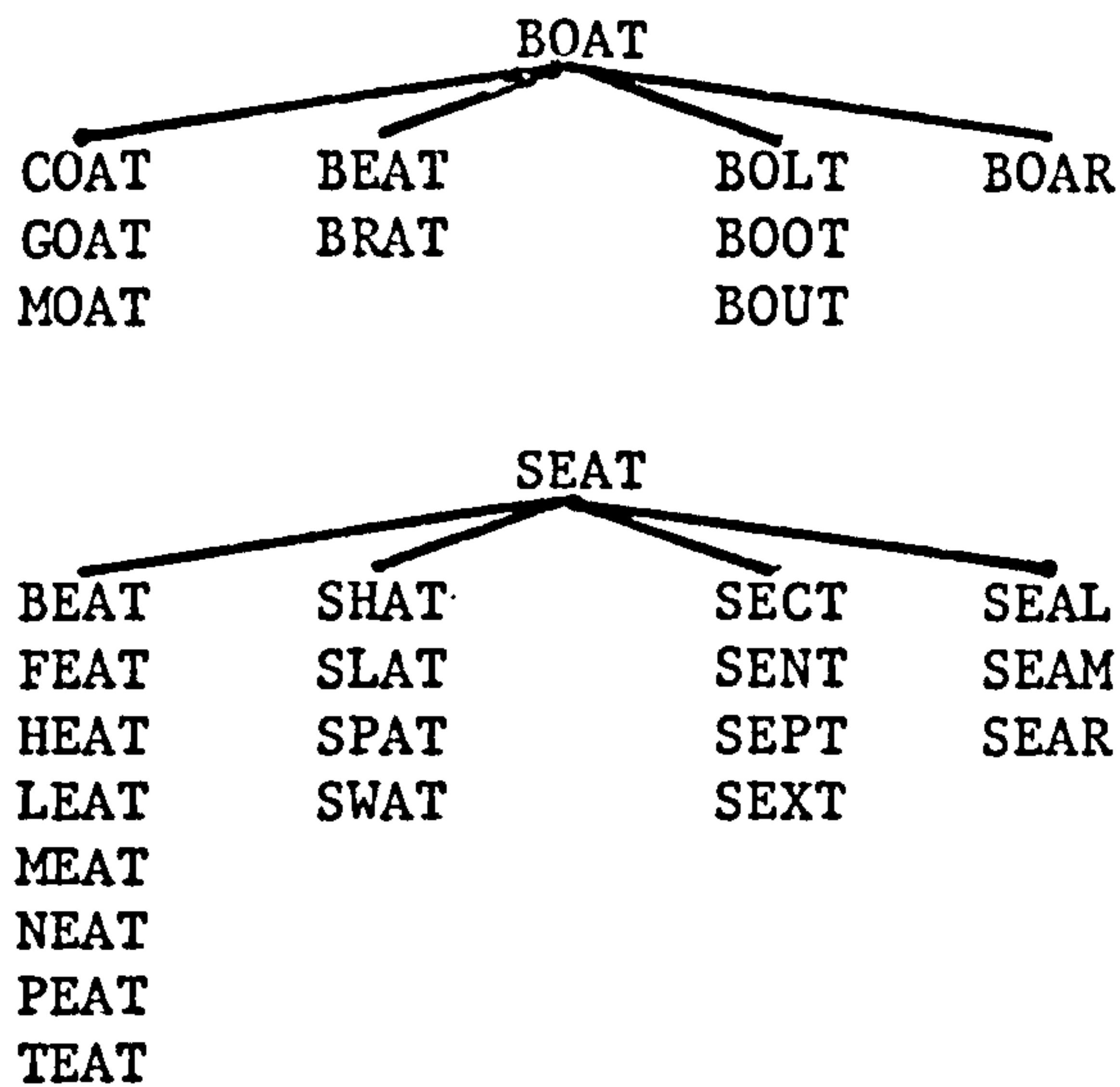
are different from other digraphs which also occur in Spanish, like rr, in that they are treated as separate letters of the alphabet, coming after c and l respectively. This means that in Spanish colmo precedes charco in alphabetical order, and luna precedes llano. More importantly, a word like chillar is technically only a five letter word, although it actually contains seven graphic symbols. One might speculate that words containing digraphs might be easier for native speakers of Spanish than words of equivalent physical length that do not contain digraphs, but this question is one that does not appear to have been investigated at all. It was thought wise to exclude words containing such forms from the material used in this experiment.

A second problem that arises from the choice of materials is that the length control may not be as satisfactory as it appears to be at first sight. The reason for this is that there is a discrepancy between the distribution of long and short words in English and Spanish. Spanish words tend on the whole to be rather longer than their English counterparts - or rather there is a complex interaction between length and frequency for these two languages. English has a higher proportion of short words among its most frequent items, whereas Spanish high frequency items tend to be longer than in English; this difference gets smaller with infrequent words. It is possible that a distributional property of this sort might allow native speakers of Spanish to develop word-processing strategies that differ quite widely from those habitually used by native English speakers when they are faced with English words. Very little is known about word processing strategies, and the sort of claims that have been generally made for English (e.g. by Fay and Cutler (1977))

are extremely crude. But as an example of the sort of problem that might arise, consider four-letter words in English. The very large number of words of this length means that it is often possible to produce a large number of other four letter words in English, which differ from an originally given word by just one letter. Figure 5-1 below shows the set of words that can be derived from BOAT and SEAT in this way.<sup>14</sup>

Figure 5-1

One-letter replacement sets of BOAT and SEAT



The fact that any of the original letters can be changed to produce a large number of other words must mean that native speakers of English process words of this type in a way that is quite different from the way they process rather longer words such as ELEPHANT.

The one-letter replacement sets of most long words are empty.

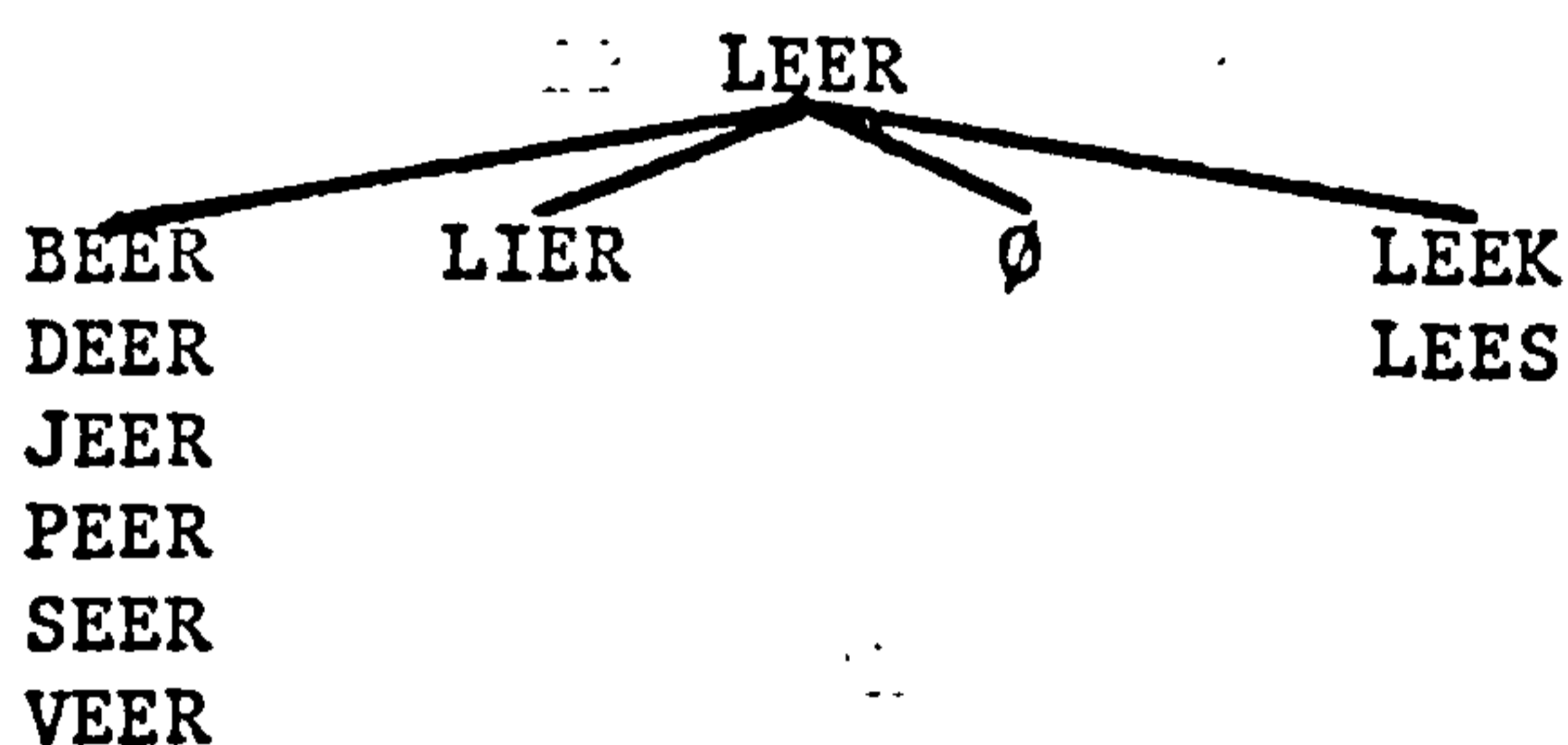
If we compare Spanish words in the same way, there is some indication that the one-letter replacement sets tend to be rather smaller than is the case for English. It is very difficult to do proper

comparisons here, since the size of the replacement sets differs widely depending on factors such as syllable structure, and the presence of consonant clusters which may not be permissible in both languages. Probably the closest match that could be found consists of sequences of letters which are acceptable words in either language, and the one-letter replacement sets of two such words LEER and MIRE will be found in Figure 5-2 below.

Figure 5-2

One-letter replacement sets of LEER and MIRE in both English and Spanish.

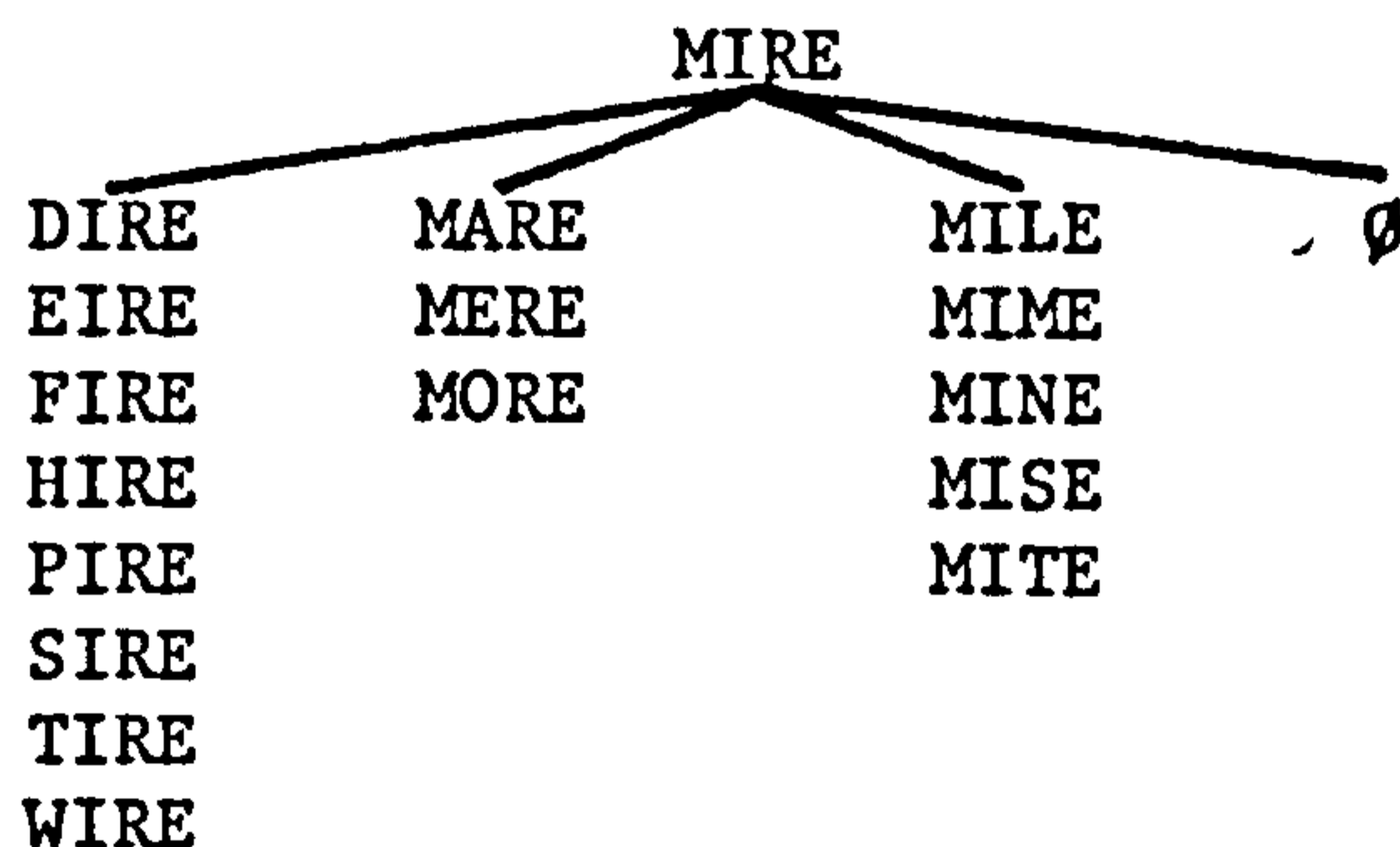
a) LEER (English)



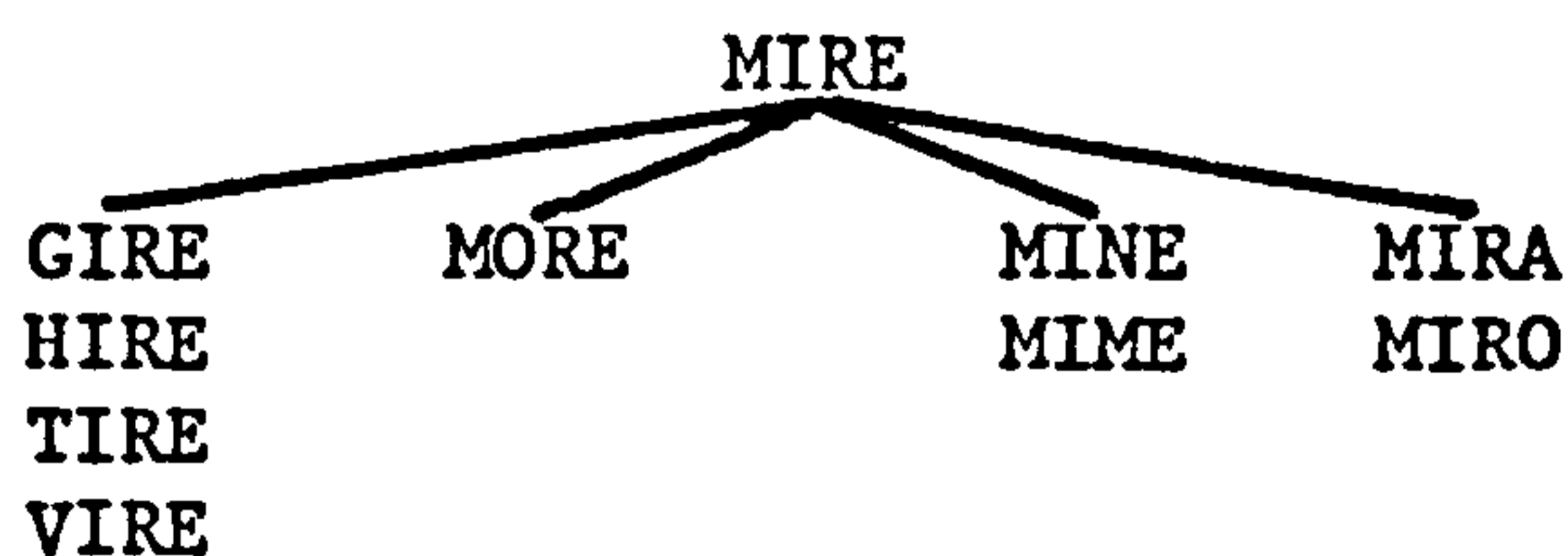
b) LEER (Spanish)



c) MIRE (English)



d) MIRE (Spanish)



It is clear from Table 5-2 that the replacement sets of the Spanish words are considerably smaller than those of the English words.

If this finding is generalizable, it would probably suggest that short words in Spanish ought to be rather more easy to handle than words of equivalent length in English on the grounds that Spanish words are more distinctive. The same argument probably holds for five letter words such as those used in this experiment, though what happens in the case of still longer words is unclear.

A further argument which reinforces the idea that the length control may not be wholly adequate derives from the morphological structure of Spanish. Consider, for example, the structure of verbs.

Citation forms of verbs in Spanish are marked with one of three infinitive endings -AR, -ER or -IR. These endings undergo changes if the verb is used in context, while the stem of the verb usually - though not always - remains unaltered. In Figure 5-2d, for example, MIRE is a form of the verb MIRAR (to look at) and the two forms in the rightmost column of this part of the figure (MIRA and MIRO) are also parts of the same verb. This raises the question of whether five-letter verbs such as MIRAR, COMER and VIVIR are actually treated as five-letter words, or if they are effectively handled as three-letter words with largely predictable inflections. A similar case can be made for nouns and adjectives to be treated as a stem plus an inflection. Noun and Adjective endings in Spanish are very limited, with a large proportion of these words ending in -O or -A. This termination serves as a singular marker, and also as a marker of gender, so that NIÑO is a boy-child, while NIÑA is a girl-child. Again it is possible here that we have a three-letter



formative NIÑ- plus a highly predictable morphological suffix, rather than a simple four-letter formative. There is some indication that words in English which contain prefixes and suffixes are decomposed into their constituent morphemes when they are perceived under tachistoscopic presentation (cf. Gibson and Guinet (1971) and Murrell and Morton (1974) for some supporting evidence). As Taft and Forster (1974) point out,

"it would seem logical and economical for the word CATS to be filed in the lexicon as CAT, and thus be recognized only after the -S has been stripped off" (p638)

and it seems equally logical and economical for MIRAR and related forms to be filed in the lexicon as MIR- and to be recognized only after the affixes have been stripped off. In practice, however, most of the work that has been done in this field has used long words, and has concentrated on the effects of affixes which affect meaning or change the form class of a formative, and not on affixes which are conditioned by grammatical and syntactic considerations, which are the ones we are interested in here. No work of this type has been carried out for Spanish.

These considerations, and others like them, suggest that perceiving five-letter words in Spanish may not be a task that is wholly comparable to perceiving five-letter words in English. This problem lies outside the scope of this study, however, as its solution would be a task of large proportions. In the meantime, given the lack of proper understanding of the processes of word recognition, we have little choice but to control for the more obvious factors such as frequency and length, and to assume that this will provide

some basis for comparability, but it is important that these caveats should be borne in mind in the discussion that follows.

### 5-3-2 Method.

The word pairs chosen were typed in upper case letters, and mounted on slides. The subjects were nine school-children studying Spanish for A-level at a London Comprehensive School. They were tested as a single group.

Each word pair was flashed onto a screen in front of the subjects for an exposure duration of one tenth of a second. Immediately after this exposure, the subjects wrote down on an answer sheet as much as they had been able to read. Two scores were taken: a TOTAL WORD SCORE, in which each word correctly reported scored one point, and a WHOLE PAIR SCORE, in which each pair of words correctly reported scored one point, but partial reports of single words scored zero.

### 5-3-3 Results.

The prediction made by the Revised Word-List Conjecture for materials of the sort used in this experiment is that in Spanish the subjects ought not to distinguish between the structured

sequences (Noun-Adjective and Noun-Verb) on the one hand, and the unstructured materials (Adjective-Adjective and Noun-Noun on the other. With the English stimuli, however, a clear distinction between the two types of stimuli would be expected, with higher scores on the structured strings.

The results of both scoring methods are shown in Table 5-4 and Figure 5-3 below. The results were submitted to an analysis of variance in which the main effects were Language and Pair-Type, and summaries of these analyses will be found in Table 5-5 and Table 5-6.

Figure 5-3. .

Mean percent total word score, and mean percent whole-pair score. Scores on the Spanish stimuli are shown hatched.

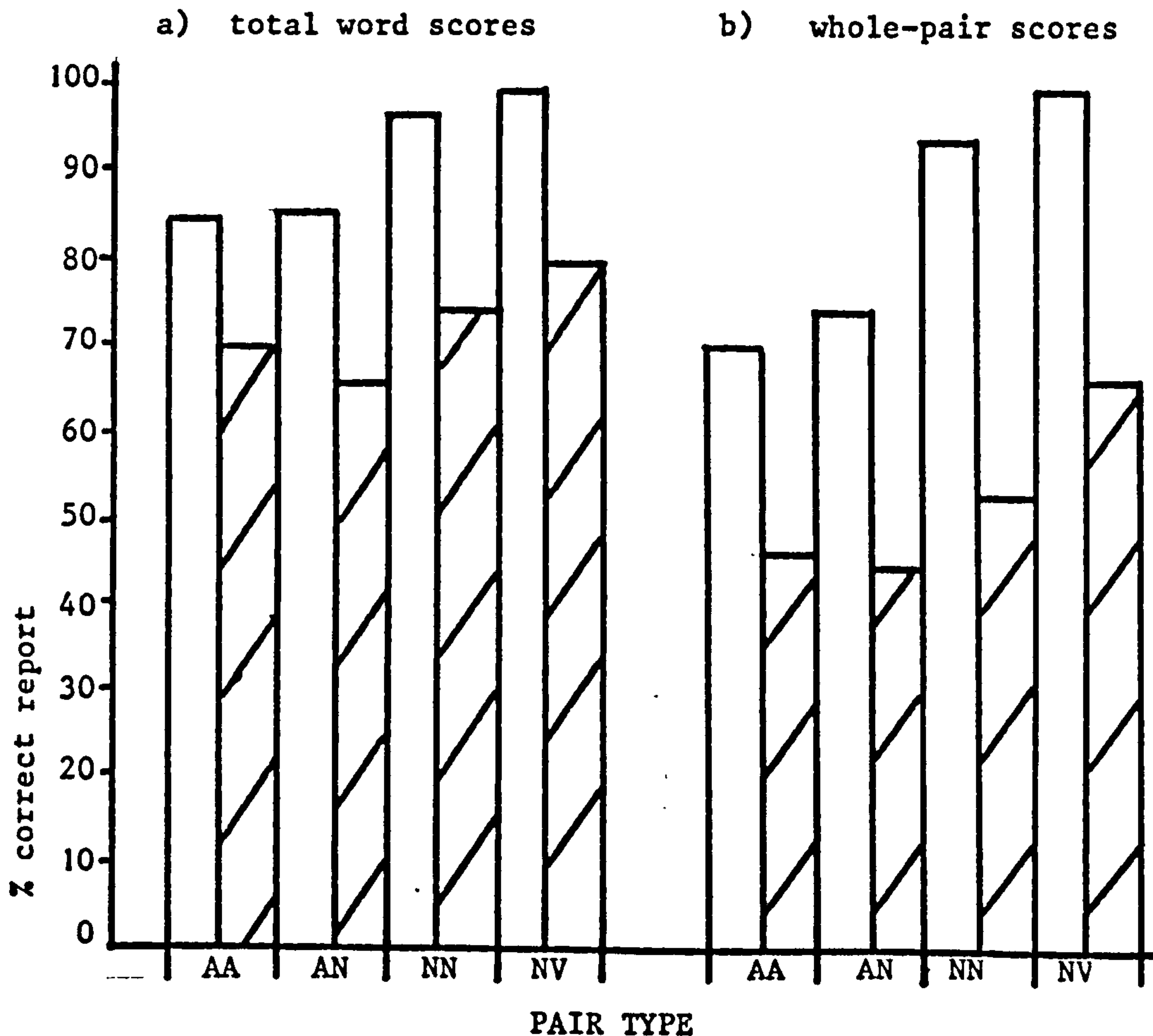


Table 5-4.

Stimuli correctly reported: Pilot Study.

a) TOTAL WORD SCORE: Max 8 words for each pair type.

PAIR TYPE:	AA	NN	AN	NV
ENGLISH STIMULI	6.78	6.89	7.78	8.00
SPANISH STIMULI	5.67	5.33	6.00	6.44

b) WHOLE PAIR SCORE: Max 4 pairs

PAIR TYPE	AA	NN	AN	NV
ENGLISH STIMULI	2.78	3.00	3.78	4.00
SPANISH STIMULI	1.89	1.78	2.11	2.67

Table 5-5.

Summary Analysis of Variance: Total Word Scores

SOURCE	df	SS	MS	F	Prob
Subjects	8	13.11			
Language	1	40.50	40.50	81.00	0.00002
Error	8	4.00	0.50		
Pair Type	3	15.33	5.11	3.64	0.027
Error	24	33.67	1.40		
Lg x PrType	3	1.06	0.35	0.48	0.696
Error	24	17.44	0.73		
Residual Error	63	112.00			

Table 5-6.

Summary Analysis of Variance: Whole Pair Scores

SOURCE	df	SS	MS	F	Prob
Subjects	8	11.75			
Language	1	29.39	29.39	53.91	0.00008
Error	8	4.63	0.54		
Pair Type	3	12.28	4.09	4.57	0.1137
Error	24	21.47	0.89		
Lg x PrType	3	1.39	0.46	1.02	0.3993
Error	24	10.86	0.45		
Residual Error	63	79.75			

For the Total Word Score, the two main effects are both significant, indicating that the subjects scored more highly in their native language than in their foreign language, and that pairs where high scores are produced in English tend also to produce high scores in Spanish. The interaction between pair-type and language is not significant, indicating that the structured word pairs produce a comparable increase in both languages.

Though the Whole Pair score is a much more sensitive measure than the Total Word Score, the results here show much the same kind of pattern as the former. Again, both the main effects are significant, and again the interaction between language and pair-type is not significant, indicating that the differences between the structured and the unstructured strings is not language sensitive.

Both these analyses seem to indicate that the subjects do make use of the syntactic structure of the foreign language strings when it is available, but three things in the data suggest that this interpretation might be an oversimplification, and should not be taken as definitive. Firstly, the severely restricted number of sequences for each pair type means that the figures are probably unreliable. Secondly, there is a clear ceiling effect for the English structured sequences. All the subjects scored 100% on the Noun-Verb sequences, and only two of the nine subjects failed to do so on the Adjective-Noun sequences. This means that there is potentially a much greater disparity between the structured sequences in English and Spanish, which might be demonstrable if a more sensitive instrument was used. Thirdly, there is a hint that the Noun-Adjective sequences in

Spanish are relatively more difficult than the corresponding Adjective-Noun sequences in English. Post-hoc analysis indicated that the difference between the Adjective-Adjective sequences and the Adjective-Noun sequences in English was significant at the 5% level ( $t=2.68$  with 8df), while the equivalent comparison for the Spanish stimuli was not significant ( $t=0.43$  with 8df). This discrepancy seemed worthwhile exploring further. A more detailed experiment was therefore carried out, and this is reported in full in the next section.

5-4. MAIN EXPERIMENT.

The object of this experiment was to test for the effects of syntactic structure on tachistoscopic recognition thresholds for word pairs of specific types in English and Spanish.

#### 5-4-1 Materials.

Two sets of materials were prepared, one in English and the other in Spanish. The English set consisted of ten pairs of adjectives and ten Adjective-Noun pairs. All the words were highly frequent (Adjective and Adjective-Adjective sections of the Thorndike-Lorge list) and were five letters long. The Spanish set also comprised ten Adjective pairs, and these were complemented by a set of ten Noun-Adjective pairs. Again the words were highly frequent; they were chosen from the vocabulary of two textbooks for beginners, and all but four of the words occurred in Juilland's count of the five thousand most frequent words in Spanish (Juilland and Chang-Rodriguez (1964)). A complete list of these stimuli will be found in Appendix 5C. Each word pair was typed in upper case letters on a plain white card.

#### 5-4-2 Subjects.

Two groups of subjects were tested: Group S, a group of ten native speakers of Spanish, temporarily resident in London, and following courses at London Language schools; Group B, a group of ten native English speakers studying Spanish, of whom six were school children studying for A-level Spanish, the remainder following similar courses in Spanish at Birkbeck College. In addition to these two groups, a third group of advanced Spanish learners (Group A) who



were all teachers of Spanish was studied. The results of this group are reported separately, in Section 5-4-5.

#### 5-4-3 Method.

Each subject was tested individually. Recognition thresholds for each word pair was established by the method of ascending limits. Each pair was exposed for 10 m.secs., and if the subject failed to report both words accurately the exposure duration was increased in increments of 10 m.secs. until a wholly correct report was produced. The strings were presented in a randomized order, and the test was preceded by four warm-up pairs.

#### 5-4-4 Results.

For each subject a mean recognition threshold for each stimulus type was calculated. In calculating this mean, responses of more than 130 m.secs. were ignored. These responses (1%) were caused by the subject's becoming fixated on an incorrect response such as lose for loose, and refusing to alter his report. Cases where subjects reported that they did not recognize one of the foreign words, or failed to understand its meaning were also left out of the counting. These responses accounted for 2% of the Group S scores, and 5% of the Group B scores.

One of the Spanish A-A sequences was subsequently deemed to be ambiguous, and on one reading could be interpreted as an Adjective Noun sequence. This item was therefore omitted from the scores of all the subjects.

The results of the experiment are shown in Table 5-7 and Figure 5-4. These results were submitted to an analysis of variance in which the main effects were Group, Language and Pair Type. A summary of this analysis is shown in Table 5-8.

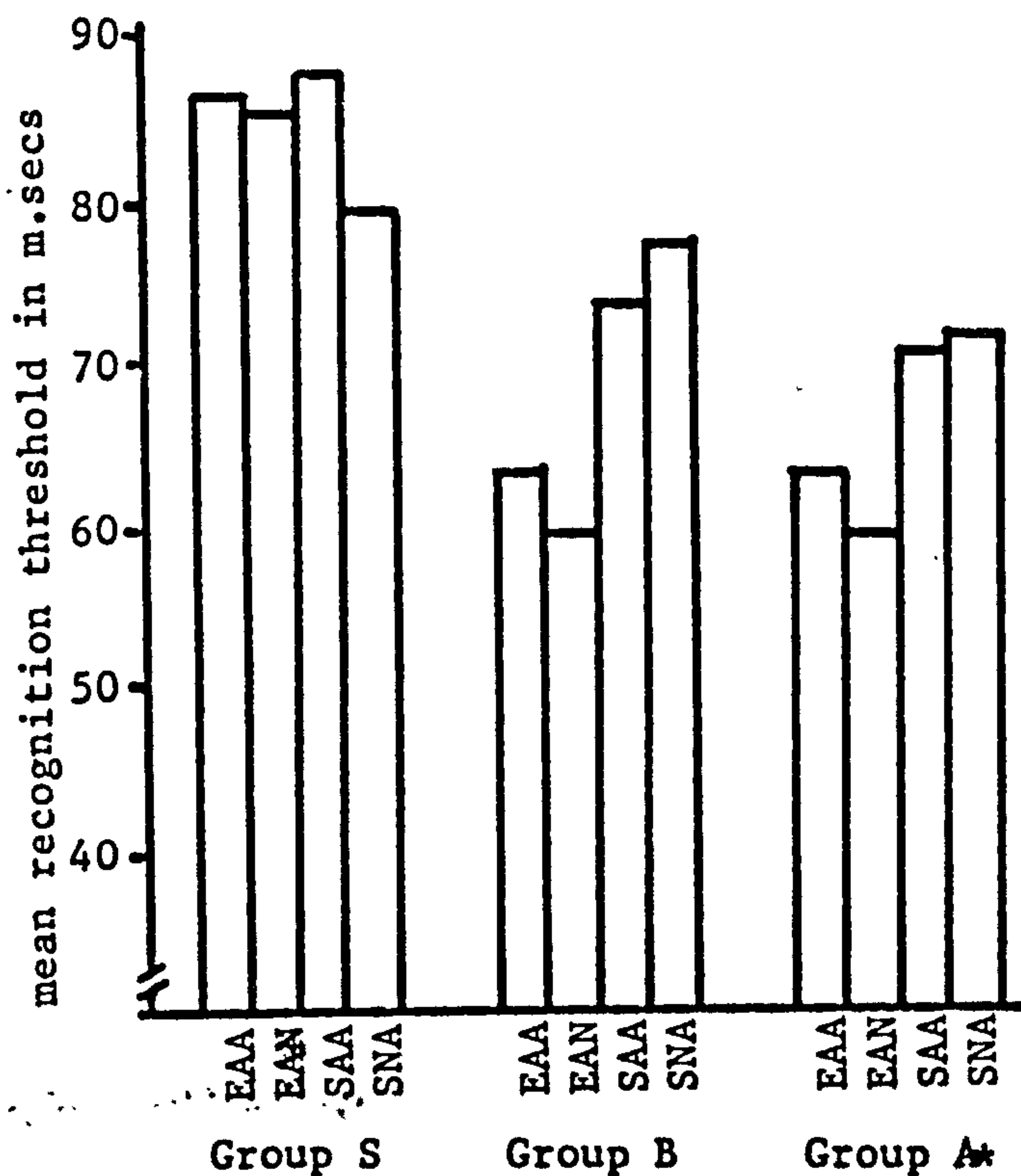
Table 5-7.

Mean recognition thresholds for structured and unstructured word pairs in English and Spanish. (in msec)

MATERIAL PAIR TYPE	ENGLISH		SPANISH	
	AA	AA	AA	NA
Group S	86.71	85.98	87.76	79.38
Group B	63.30	59.80	73.55	77.23

Figure 5-4.

Mean recognition thresholds for structured and unstructured word pairs in English and Spanish. (in msec.)



\*see section 5-4-5

Table 5-8.

## Summary Analysis of Variance.

SOURCE	df	SS	MS	F	PROB
Subjects	19	11960.95			
Group	1	5436.75	5436.75	14.99	0.0011
Error	18	6524.20	362.46		
Language	1	612.17	612.17	14.67	0.0012
Lg x Group	1	1380.29	1380.29	33.08	0.00002
Error	18	751.04	41.72		
Pair Type	1	99.68	99.68	2.36	0.142
Pair Type x GP	1	107.88	107.88	2.55	0.128
Error	18	761.01	42.28		
Lg x Pair Type	1	0.28	0.28	0.02	0.897
Lg x GP x PType	1	274.91	274.91	17.16	0.00061
Error	18	288.45	16.03		
Residual Error	60	4275.71			

This analysis showed that two of the main effects were highly significant. The English native speakers, Group B, produce overall lower thresholds than the native speakers of Spanish, Group B. Overall the English stimuli are reported at lower thresholds than the Spanish stimuli. There is also a highly significant interaction between language and Group and a further highly significant interaction between the three main variables.

The least important of these is the significant Group effect. The most probable explanation of this finding is that the native Spanish speakers were on the whole older than the native English speakers, who were predominantly under twenty-five. This latter group might be expected to have sharper eyesight, even though all the subjects had normal or corrected vision.<sup>15</sup>

(Similarly high thresholds for Spanish words perceived by native Spanish speakers can be inferred from an experiment reported by Walters and Zatorre (1978). In a test with a single exposure of two unrelated words and a digit for 40 m.secs., Walters and Zatorre found that native English speakers performed better than native Spanish speakers in either language. They suggest that the native English speakers, who were all students of Spanish, might "have been more practiced in reading and word identification as part of their daily studies than the (Spanish) students in the other group, who were more heterogeneous with respect to field of study. Differences in mode of learning may also account at least partially for these group differences. It can be assumed that the native English group learned Spanish through a combination of the visual and the auditory modalities. The native Spanish speakers are more likely to place greater reliance on auditory and verbal strategies in English than the other group, since many linguistic interactions upon coming to the United States were necessarily verbal in nature." (p164) This does not seem to be a wholly satisfactory explanation of why a group of non-native speakers should have higher performance levels than a group of well-educated native speakers, and the fact that two independent studies have both come up with the same unexpected finding suggests that there may be something worth investigating here.)

The significant language effect and the interaction between group and language appear to be due to the fact that the native English speakers have considerably lower thresholds for English material than they do for Spanish. The native Spanish speakers, Group S, do not show

this difference, nor do they show any clear advantage for Spanish words over English words. There is no obvious explanation for these findings, though it is possible that what is emerging here is signs of different processing strategies of the sort mentioned in 5-3.

Since the three-way interaction between Language, Pair Type and Group was also highly significant, separate analyses of variance for each group were carried out. The main effects in these analyses were language and Pair Type, and a summary of the tests will be found in Table 5-9 and Table 5-10.

Table 5-9

Summary of Analysis of Variance: Group S.

SOURCE	df	SS	MS	F	PROB
Subjects	9	5532.45			
Language	1	77.01	77.01	1.47	0.2555
Error	9	470.01	52.22		
Pair Type	1	207.48	207.48	3.31	0.1024
Error	9	564.94	62.77		
Lg x PType	1	146.31	146.31	8.90	0.0154
Error	9	147.89	16.43		
Residual Error	30	1613.63			

Table 5-10

Summary of Analysis of Variance: Group B.

SOURCE	df	SS	MS	F	PROB
Subjects	9	991.75			
Language	1	1915.46	1915.46	61.34	0.00003
Error	9	281.03	31.23		
Pair Type	1	0.08	0.08	0.0037	0.9526
Error	9	196.06	21.78		
Lg x PType	1	128.88	128.88	8.25	0.0184
Error	9	140.56	15.62		
Residual Error	30	2662.08			

These analyses show that despite other differences, the two groups are similar in that they both show a significant interaction between language and pair type. In both cases this interaction is caused by the occurrence of relatively low thresholds on the structured material in the native language, and the absence of such low thresholds for the structured sequences in the foreign language. This result offers clear support for the Revised Word-List Conjecture.

#### 5-4-5 Supplementary Experiment.

The main obstacle to accepting this result at face value is the large difference between the languages for the native English Group. One possible explanation for this difference is that the members of this group were relatively less proficient in Spanish than the native speakers of Spanish were in English. This suggestion is supported by the fact that more errors and non-responses were recorded with Group B than with Group S. The experiment was therefore repeated with a second group of native English Speakers. This group were very advanced learners of Spanish; all of them had degrees in Spanish, and used the language in their jobs, (nine of these S's were language teachers, the other S was a bilingual secretary). The results of this group (Group A) are shown in Table 5-11, and Figure 5-3 above.

Table 5-11.

Mean recognition thresholds for structured and unstructured stimuli in English and Spanish: Group A. (in msec.)

MATERIAL	ENGLISH		SPANISH	
	AA	AN	AA	NA
Group A	63.10	59.20	70.34	71.25

Surprisingly, the more advanced non-native group does not show any great difference from the less advanced group. The Spanish sequences still produce significantly higher thresholds than the English ones, ( $F=52.8$ ,  $p<.001$  with 1,9 df) and there is a significant difference due to pair type in English, but not in Spanish ( $t=2.72$ ,  $p<.05$  and  $t=1.01$ ,  $p<.05$  with 9 df respectively.)

## 5-5. DISCUSSION.

The results of the experiment reported in the previous section offer clear support for our Revised Word-List Conjecture. Native speakers of both English and Spanish show lower recognition thresholds for structured material in their native language, but not for comparable material in their foreign language. Structured word pairs in the foreign language are indistinguishable from unstructured word pairs, and this finding is exactly in line with predictions derived from the Word-List Conjecture.

As usual, however, there are a number of reservations that need to be made at this stage.

The first reservation concerns the length of the strings tested. All the material used in the experiment consisted of pairs of words, i.e., of relatively short strings, and it is not clear what sort of results would have been found if longer sequences had been used. It seems reasonable to infer that longer sequences would have produced essentially similar results, however, since learners would be unlikely to use syntactic structure in long sequences if they were unable to do so in shorter sequences. Clearly, however, it would be worthwhile to test this idea further using rather longer stimulus strings.

The second reservation concerns the type of structure used in the experiment. Only one type of structure figured in this test -



word pairs consisting of a Noun and an Adjective which could be interpreted as forming a Noun Phrase. It might be argued that the lack of a structure effect reported here would not be likely to extend itself to other types of structured material, but that this Noun Phrase structure is exceptional in some way. As we have seen, this particular structure is one where there is a marked difference in English and Spanish. English usually uses preposed adjectives, while Spanish prefers postposed adjectives as a general rule. This difference in word order is a fairly gross distinguishing characteristic, and it might be reasonably supposed that major word order differences of this sort between languages would be a source of some difficulty to non-native speakers, since they involve radical alterations in the transitional probability patterns between words. It is possible, therefore, that the difference in performance which we have ascribed to the ability of subjects to react to syntactic structure in strings of words may be a special case of a contrasting structure, rather than an illustration of a more general lack of sensitivity to syntactic structure on the part of learners. Had we used, say, structured pairs of words consisting of Verbs and Nouns chosen to make up plausible Verb Phrases - where there is typically no contrast in English and Spanish - or if we had chosen to study structures that contrasted in some more subtle way, then it is possible that findings in support of the Revised Word-List Conjecture would not have materialized.

This line of argument obviously owes much to the ideas of Contrastive Analysis -cf. for instance, Stockwell, Bown and Martin (1965) or for a more recent reworking in terms of Universal Grammar,

cr. Aronson Berman (1978). It represents a significant development of the traditional contrastive approach, however, since Contrastive Analysis generally limits itself to predicting and accounting for only the errors made by second language speakers (cf. Schachter (1974)), and for this reason it has been one of the chief tools of the Error Analysts whose work was criticized in Section 1-2. We argued there that such approaches were unduly restricted in scope, and suggested that there was a strong case for regarding errors, whether in reception or production, as merely extreme points on a continuum of non-native speaker performance, rather than as an isolated set of phenomena, to be treated wholly independently of other aspects of learners' behaviour in a second language. Even when structures are correctly handled - in the sense that they are accurately formulated or properly understood - it is still possible that some difficulty of a less overt kind might be experienced by an L<sub>2</sub> speaker if psychologically important contrasts were involved. Given that recognition thresholds are a much more sensitive tool than a catalogue of errors, it might be that the results are picking up difficulties of this sort, rather than being caused by any real lack of sensitivity to syntactic structure.

The results of this chapter are certainly compatible with a contrastive explanation along these lines, and indeed, the indications from the Pilot Study (5-3) which showed that Verb-Noun sequences were perceived considerably better than Noun-Adjective sequences or unrelated word pairs in Spanish, would tend to support this analysis. On the other hand, if contrasting structures were

all that was involved, then one might expect the native Spanish speakers to have been rather less affected by the English structures than the native English speakers were by the Spanish structures. Since Spanish does allow preposed adjectives to occur as well as the more common postposed ones, and since this pattern occurs fairly frequently, it could be argued that native speakers of Spanish ought to have developed a strategy that would enable them to handle ( $_{NP}(\text{Adj N})$ ) sequences with relative ease, and they should show some signs of lower thresholds for the Noun Phrases in English than for the unstructured strings. The results do not support this claim, however, and this suggests that the contrastive explanation of the effect is only partially correct.

The Contrastive explanation represents a very serious limitation on the Revised Word-List Conjecture, and in view of this it seemed unwise to reject it out of hand on the basis of the rather flimsy evidence reported above. The obvious next step, therefore is to undertake a series of further experiments using the tachistoscopic recognition threshold paradigm, and in particular to use this method as a way of exploring the behaviour of non-native speakers when they are faced with structures which are identical in two languages. A number of practical considerations militated against this obvious course of action, however - namely, some limitations which seem to be inherent in tachistoscopic work, and some other problems which arise out of the method of ascending limits in particular. These considerations are discussed in detail below.

The principal problem arises from limitations on the length of string which can be used in threshold experiments. Because of the way the eye is constructed, strings of letters which subtend a visual angle of more than ten degrees cannot be apprehended in a single glance, and this limit, combined with the physical properties of tachistoscopes, imposes a severe restriction on the length of string that can be used in an experiment of this type. This may seem unimportant, since, as we have seen, it is possible to study a range of constructions consisting of two words, which would lie within this limit, but the use of two-word strings is not wholly satisfactory in this context. Firstly, there are instances of many syntactic types occurring fairly frequently, which cannot be adequately conveyed in a two word string - Indirect Object constructions are one obvious example of this, Subjects of Transitive Verbs are another. In general, while we limit ourselves to two-word stimulus strings, it is impossible to study the effects of the higher-level syntactic structures with any thoroughness, simply because three or more formatives are required before we begin to find instances of hierarchical structuring of this more complex variety. Indeed, even in the study of Noun Phrases such as were used in the experiment reported in the previous section, two words is not really sufficient, since our strings consisted of Nouns and Adjectives, but not Determiners which are technically necessary for a complete Noun Phrase structure. We have been referring to SHARP KNIFE or VIAJE LARGO as Noun Phrases, but this to some extent is an oversimplification imposed by the paradigm, since they are actually only truncated Noun Phrases. Other

difficulties arise if we attempt to use any of the other obvious word pairs in a similar design. For instance, if we use (<sub>S</sub> NV) sequences such as DOG BARKS, then it could be argued that the missing determiner makes this a very artificial task, possibly something more akin to a word association task than a study which tests syntactic structure. Similarly, if we use word pairs which illustrate a (<sub>VP</sub> VN) structure, such as BRUSH TEETH, it could be argued that the missing Determiner is problematical, and even if this is avoided by using mass nouns as in SPENDS MONEY, then the effects of missing Subject needs still to be taken into consideration. Interestingly, this latter problem is not one that arises in Spanish if we use finite verb forms, since these verbs contain their own subjects in the form of a morphological inflection. Thus GASTA DINERO can stand on its own as a complete - albeit somewhat cryptic - sentence (= he spends money) in a way that SPENDS MONEY cannot. This means that comparisons of Verb-Noun structures in English and Spanish is much more difficult than it appears at first sight; it is not actually possible to isolate a (<sub>VP</sub> VN) structure in Spanish, since any structure of this type would automatically be interpreted as a (<sub>S</sub> VN) structure - a form which does not occur in English except in Imperative sentences.

A further criticism that could be made of the work reported in this chapter is that the method of ascending limits is itself considered by some scholars to be an unsatisfactory procedure. The successive exposures of the same material allow the subject to guess the second word on the basis of the first one he manages

to recognize, and this means that the threshold figures obtained are not true thresholds, but only reflections of the subjects' guessing ability. According to this criticism, the results reported here show only that the subjects do not guess, or do not guess so accurately, in their foreign language. This criticism is probably less serious than it appears at first sight, however. We know that guessing and predictive behaviour in general plays a central role in our ability to process languages, and that syntactic structure provides important clues that facilitate this guessing. If learners could be shown to be systematically ignoring these clues, even in a situation that actively encourages them to guess, then this would obviously amount to an important difference between them and native speakers.

In theory, one way of avoiding these criticisms of the method of ascending limits is to use single presentations, as in the pilot study, but in practice, there are a number of difficulties in using this latter technique in an experiment that compares native speakers and learners of a language. In order to get reliable data, it is necessary to use a large number of word pairs at an exposure duration that does not allow all the material to be read, but which is not so short that none of the material is readable. Both of these factors cause difficulties for experiments with second languages. Firstly, it is important to use only high frequency words which the learner groups might be expected to know, and this severely limits the number of usable sequences, especially if additional restrictions such as length constraints are also invoked.

In the main experiment reported in Section 5-4, ten five-letter word pairs were used for each type of structure, but it would have been very difficult to produce a number much larger than this within the constraints adopted, and almost impossible to find a sufficient number of acceptable stimuli to make a single exposure experiment a practical proposition. Secondly, given that learners have much higher thresholds for foreign language words than for words in their native language it is difficult to see how a single exposure could work effectively. An exposure duration of the right length for native language words would be rather too short for foreign language words, while an exposure duration of the right length for foreign language words would be far too easy for native language words. This makes it difficult to carry out within subject comparisons of the sort used in the experiment reported above. The results of this experiment suggested that individuals' abilities to recognize words varied considerably, and though there was a reliably higher threshold for foreign language words, it would be a task of some difficulty to choose an arbitrary exposure duration which would produce the right level of successes and failures in a wide range of subjects.

Considerations of this type strongly suggested that further work using tachistoscopic recognition thresholds might not be the best way of pursuing this investigation, and accordingly another experimental technique was sought which could be used to throw light on the structuring behaviour of non-native speakers.

One obvious candidate for this is a technique for assessing rapid word processing which is rather more sophisticated than the recognition threshold method used in this Chapter - the Rapid Serial Visual Presentation (RSVP) technique devised by Forster (1970). In this method, subjects are presented with word lists or sentences one word at a time on successive frames of a film strip, and asked to read them aloud. The words are presented at a rate which is so fast that a high proportion of errors occur, and the main point of interest is the immediate causes and position of these errors. This technique has a number of advantages over the more traditional tachistoscopic presentation. Firstly, since all the words appear on the screen in the same position, the time for which each part of the stimulus string is available for inspection can be closely controlled. This is impossible when all the words in the test string are presented simultaneously. Secondly, since the words appear one after the other, rather than side by side, much longer sequences can be used than in the normal tachistoscope experiment. Forster's method has been used to investigate the effects of clause structure on perception with some success, and it seems that the technique could also be used with relatively minor alterations to assess the effects of within clause structure too, although work of this type does not appear to have been undertaken. In principle, RSVP ought to be a very useful tool for investigating the performance of non-native speakers with a wide variety of test materials, and it is an obvious candidate for investigating the effects of syntactic



structure on non-native speakers. In practice, however, it turned out that the technique required a level of technical support which was not readily available at the time this research was carried out, and consequently, RSVP was not used. Instead a further structure-sensitive task - the Probe-Latency technique devised by Suci, Ammon and Gamlin (1967) - was taken up and a series of experiments that exploit this technique are reported in the next chapter.

**CHAPTER 6.      PROBE LATENCIES**

**6-1    Introduction**

**6-2    Background**

**6-3    Extensions and Developments of the Probe Latency Paradigm**

**6-4    Probe Latencies and Second Language Speakers**

**6-5    Experiment 1**

**6-6    Experiment 2**

**6-7    Discussion**

## 6-1. INTRODUCTION.

In Chapter 5, some evidence, derived from a study of Recognition Thresholds for tachistoscopically presented phrases, was presented. This evidence supported the Revised Word-List Conjecture - the claim that non-native speakers are insensitive to syntactic structuring in foreign language sentences except when this is signalled by high frequency grammatical function words. It was argued, however, that this data was also compatible with an alternative claim; since the Noun Phrase structures studied were instances of structures where there is a gross contrast of word-order in English and Spanish, it is possible that the Revised Word-List Conjecture may only apply to contrasting structures, and that where structures do not contrast in this way, it is possible that non-native speakers may be able to use processing strategies developed for their native language to enable them to perform in a manner that is broadly comparable with the way native speakers perform. This claim is clearly much weaker than the Revised Word-List Conjecture.

In this chapter a series of experiments which investigate this contrastive account are reported. A new experimental technique - the Probe Latency Technique devised by Suci, Ammon and Gamlin (1967) has been adopted here, for the reasons discussed in Section 5-5, namely that this technique allows us to study longer and more complex stimulus strings than those used in Chapter 5, some of which

contrast in English and Spanish, together with others that do not. This should allow us to ascertain whether the Revised Word-List Conjecture should be restrained along the lines indicated by the Contrasting Structures argument, or whether it accurately predicts the behaviour of non-native speakers as it stands.

The Probe Latency Technique itself, and some further work based on the original experiments of Suci Ammon and Gamlin are discussed in the next section.

## 6-2. BACKGROUND.

### 6-2-1 Suci, Ammon and Gamlin (1967)

The Probe Latency Technique was first used by Suci, Ammon and Gamlin (1967) as a way of "assessing how language is structured by a hearer" (p69). The technique works as follows. The subject listens to a sentence and immediately following this, he hears a probe word, i.e. an isolated word that was part of the preceding sentence. The subject's task is to respond as quickly as possible with the word in the sentence that followed the probe word. For example if the test sentence were:

the three men walked into the shop

and the probe word man, then the subject's task would be to respond as quickly as possible with the word walked. Using the same example, if the probe word were into then the subject should respond with the. The dependent variable measured is the latency of this response.

Suci, Ammon and Gamlin's original experiment used two sentence types. Type I sentences consisted of the sequence Article, Adjective, Noun, Verb, Adjective, Noun. The Type II sentences consisted of the sequence Article, Adjective, Adjective, Noun, Verb, Noun. Examples of these two sentence types are shown in Figure 6-1, together with the constituent analysis for the sentences proposed by Suci, Ammon and Gamlin.

Figure 6-1.

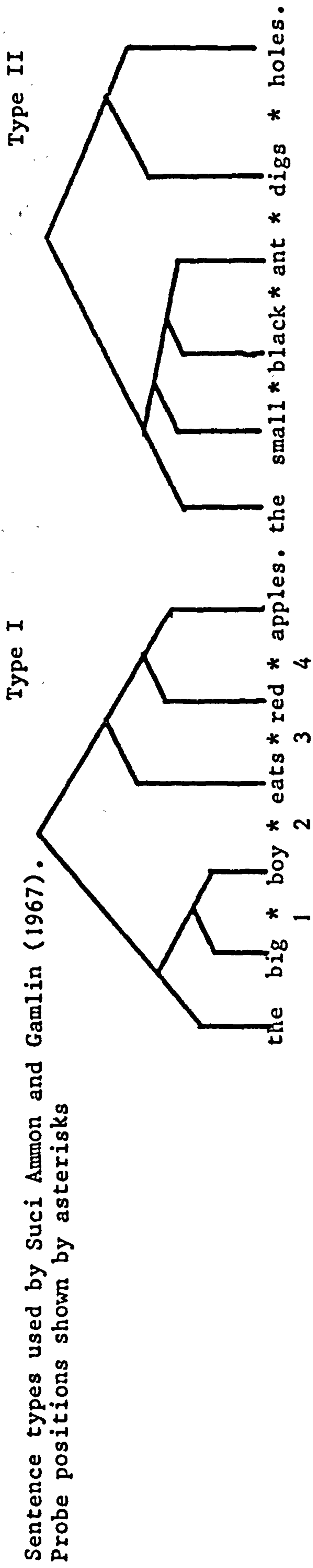
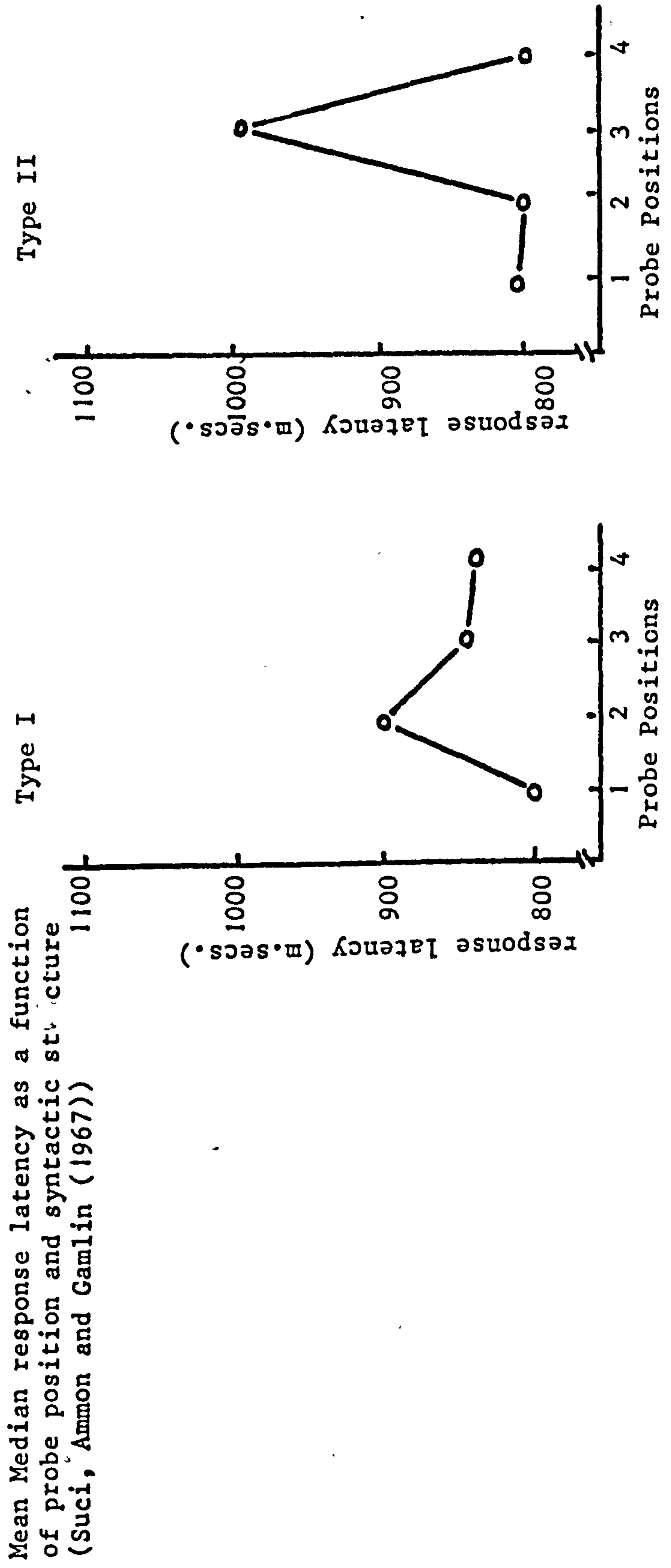


Figure 6-2.



Each of these sentence types was probed in four places, indicated by the asterisks in Figure 6-1. Each probe position was tested four times, a different sentence being used for each probe test. The results of this experiment are shown in Figure 6-2. In these diagrams, each point represents the mean median latency score of the correct responses. The two sentence patterns produce significantly different results, which can be explained in terms of their syntactic structure. In both cases, the longest latency is associated with the major constituent boundary, in that when the probe and the response straddle this boundary, the response latency is longer than when both probe and response word are part of the same constituent. Suci, Ammon and Gamlin also claim that there is some evidence for the technique being able to distinguish constituents within the Verb Phrase of the Type I sentences, in that the adjective and the noun are separated by a shorter latency than are the verb and the adjective. No statistical support is offered for this last claim, however.

#### 6-2-2 Length Effects.

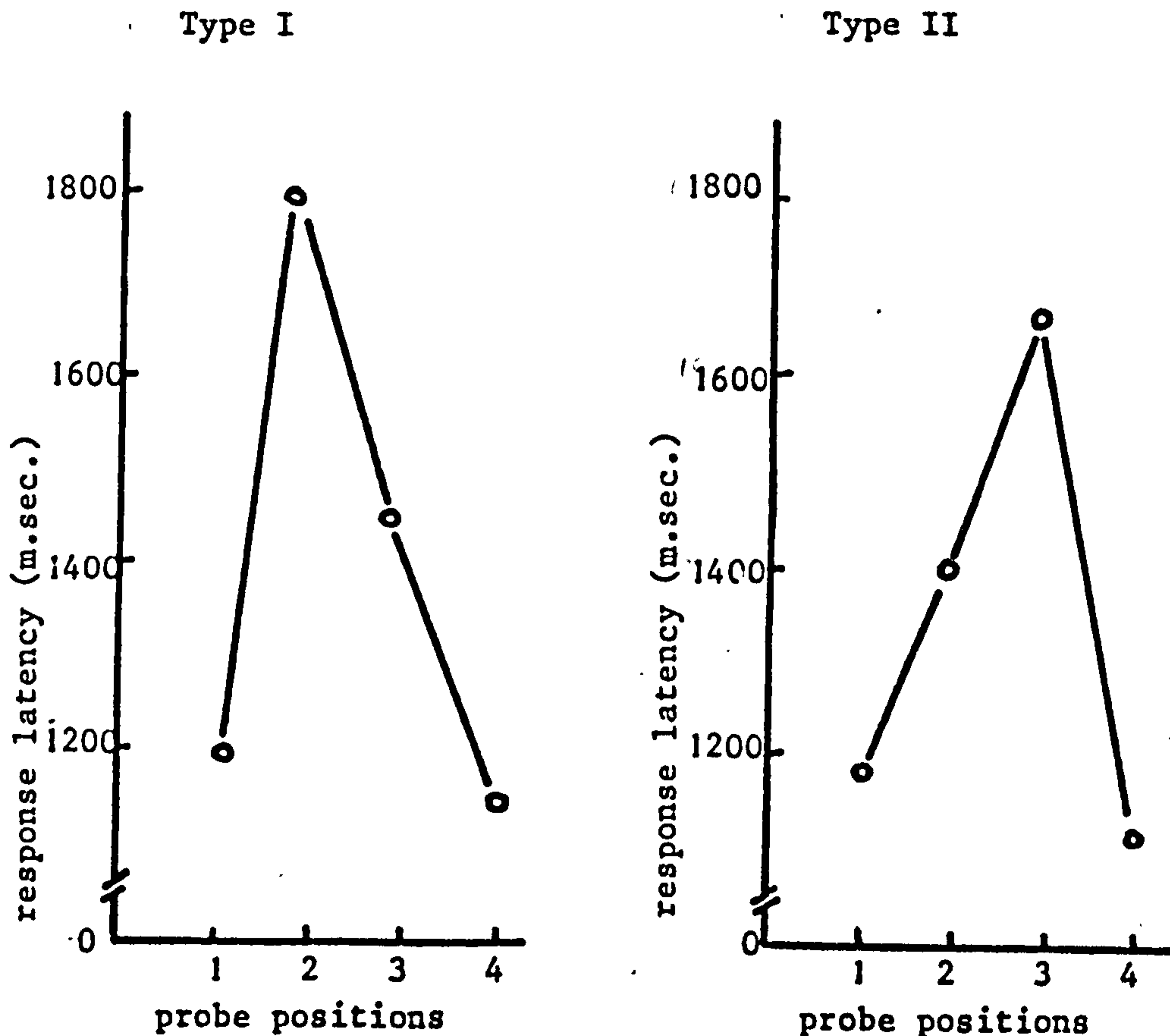
In a further experiment, Suci, Ammon and Gamlin used identical stimulus sentences but lengthened by the addition of a three-word prepositional phrase. These stimuli produced essentially similar results. Overall longer latencies resulted, but no probe position was affected more than any other.

### 6-2-3. Results with Children

Suci, Ammon and Gamlin also report that essentially similar results are found with young children, and their findings with these subjects are shown in Figure 6-3 below. The main difference between these results and those of the adults is the much longer latencies, and the much more clearly marked structural effect. Increasing the length of the sentences increases the latencies in much the same way as for adults. The children's increase is larger than that found with adults, but the overall pattern remains unchanged. The children's results show a very clear relationship between constituent structure and latency at different probe positions.

Figure 6-3.

Mean median Response latencies as a function of probe position and sentence type: Children's data (Suci et al (1967)).





#### 6-2-4 Meaningless Material

Suci, Ammon and Gamlin's final findings deal with probe latencies produced by semantically anomalous material - strings which preserve the constituent structure of English, but violate the normal selection restrictions of English. The results from this part of Suci, Ammon and Gamlin's paper are difficult to interpret. Though the same sort of differences are found here as were found for the meaningful material, the differences between probe positions are significant only in the case of the Type I sentence pattern: (Art Adj N) (V (Adj N)). Suci, Ammon and Gamlin comment "that with anomalous input, some syntactic types such as Type II, are not reliably effective in producing structure" (p79) but this is not a convincing explanation of the lack of pattern, since it merely restates the problem.

#### 6-2-5 Conclusions.

Suci, Ammon and Gamlin conclude rather cautiously that "although the relations between structure and the two other variables (meaningfulness and length) are not totally clear, it seems that the probe latency technique itself is clearly a potentially useful tool in the study of language processing. It seems to be sensitive to syntactic variations and also to individual differences in how an input is structured under different conditions. As such it deserves further attention."

This "further attention" was rapidly forthcoming, and though some of this work deviates markedly from the classic simplicity of Suci Ammon and Gamlin's study, the broad line of development supports their interpretation of the results as showing that latency is affected by syntactic structure. These later developments of the paradigm are discussed in the next section.

### 6-3. EXTENSIONS AND DEVELOPMENTS OF THE PROBE LATENCY PARADIGM.

#### 6-3-1. Summary.

There is only a small body of literature that discusses further experimental developments of the probe latency paradigm. On the whole, the results of these studies tend to confirm the earlier findings of Suci, Ammon and Gamlin, though some of these original findings appear to be more robust than others. The major constituent effect, in particular, seems to be a fairly reliable finding, though the clear relationship found by Suci et al between the fine detail of syntactic structure and latency in the data from children (cf Figure 6-3) seems to be more elusive, and rather more difficult to establish for adults.

Three main groups of studies exist in which the probe latency paradigm has been used: those of Kennedy and Wilkes, Kempen and Ammon. This work is discussed in detail in the sections that follow.

A. Kennedy and Wilkes (1968)

Kennedy and Wilkes' first experiment used a modified form of the probe latency task. In this modified version each sentence tested is repeated twice to the subject, who then repeats the whole sentence aloud. Next, each word in the sentence is probed in a random sequence at six second intervals. The subject is required to respond with the word that followed the probe in the stimulus sentence. It is not clear whether this method of probing each word in a single sentence, rather than probing one word in a large number of sentences represents an important modification to the technique or not. It clearly has a number of practical advantages. The main disadvantage, however, is that whereas Suci, Ammon and Gamlin's method is obviously a short-term memory task, Kennedy and Wilkes' method is more complex, and further removed from the actual perception of sentences. This objection may not be well-founded, however, as even in Suci, Ammon and Gamlin's original experiment there is a slight pause between the end of the sentence and the presentation of the probe word, which subjects could use for rehearsal. When Kennedy and Wilkes require their subjects to repeat the stimulus sentence aloud before the presentation of the probes, they may merely be making this rehearsal overt instead of covert. Kennedy and Wilkes' results are similar enough to Suci, Ammon and Gamlin's findings to support this interpretation.

Six sentences, all composed of monosyllabic words, and all following the same syntactic pattern were used. This pattern is

shown in Figure 6-4, and results of the experiment are shown in Figure 6-5.

Figure 6-4.

Sentence Type used by Kennedy and Wilkes (1968).

The constituent structure is that proposed by Kennedy and Wilkes

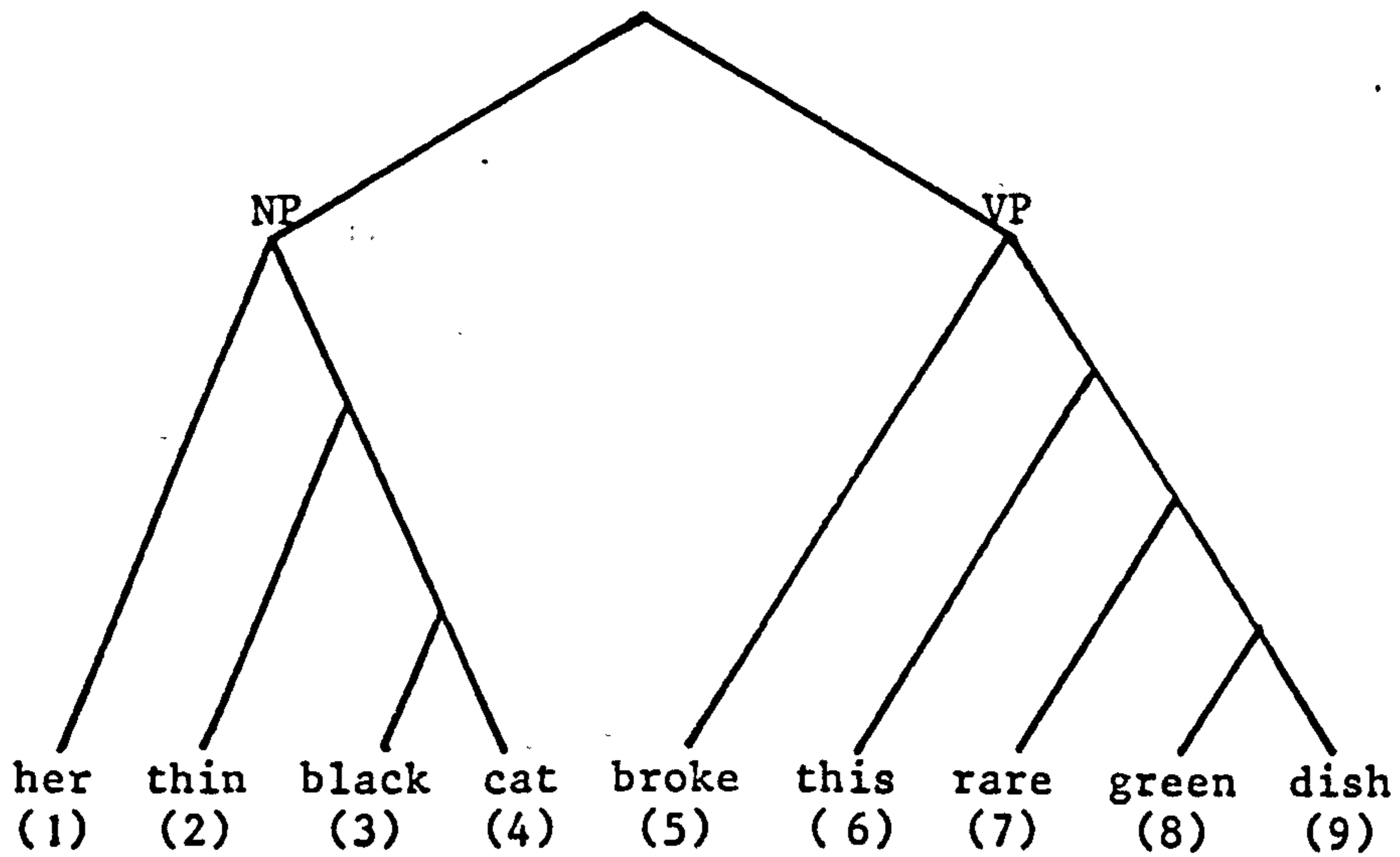
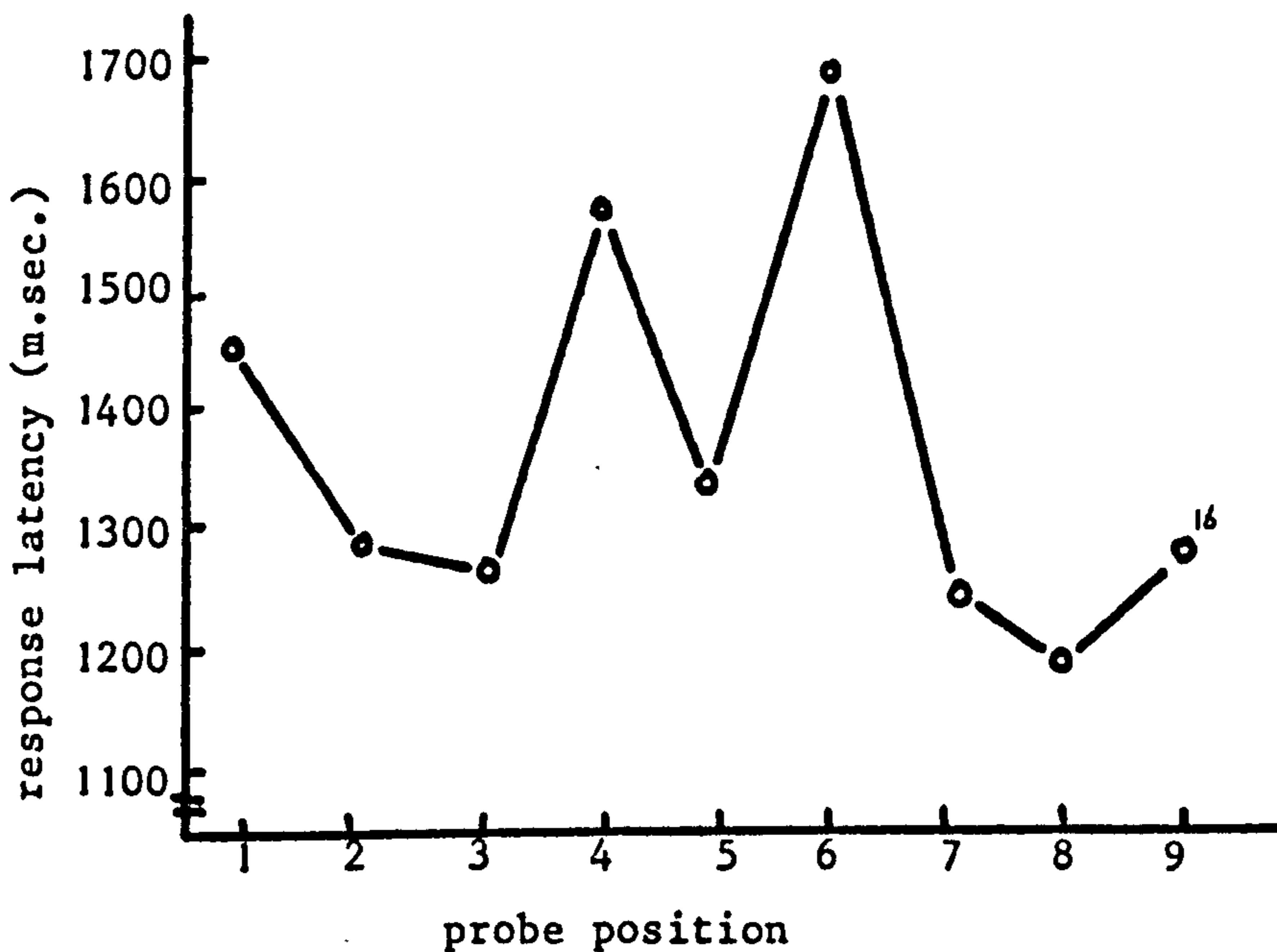


Figure 6-5.

Mean response latency as a function of serial position of the stimulus item. (Kennedy and Wilkes (1968)).



The sentences used by Kennedy and Wilkes are rather longer than Suci, Ammon and Gamlin's and the resulting latency curves are correspondingly more complex. Kennedy and Wilkes report a significant probe position effect, and more particularly, that probe position 6 produced longer latencies than all other positions except 4; that 4 and 6 exceeded all others except 1; and that 4, 6 and 1 exceeded all but 5. (Numbers refer to the serial position of the probe word, not of the response).

Kennedy and Wilkes' discussion of their findings is not primarily concerned with the relationship between probe latency and structure, since their data was collected as a test of Sternberg's (1967) theory of serial scanning. Nevertheless, Kennedy and Wilkes do raise the idea that short latency may reflect a point of "easy access" to the stored representation of a sentence, and point out that in this case, position 5, the point at which the verb phrase constituent begins would be an obvious candidate for an easy access point of this sort. "Long latency at position 4 contrasts with the short latency found at position 5, and it could be that ready access to the sentence may be made at position 5 in the scanning process. This could imply separate access to the two major constituents of this sentence type." (p393). Kennedy and Wilkes are (rightly) cautious about this idea, and it is not developed further in this paper. Later experiments were undertaken explicitly to test this hypothesis, and they are discussed in detail below. Kennedy and Wilkes point out, however, that the long latencies found at positions 1 and 6 are embarrassing for the hypothesis and suggest that they are probably caused by the

repetition of words like the and this across sentences being a source of uncertainty in the subjects. They argue that if these two scores are notionally reduced in order to make allowances for possible confusion, then the longest latency occurs across the major constituent boundary. With this correction, the results offer some support for the view that probe latencies are affected by syntactic structure. Kennedy and Wilkes fail to point out that the Heads of Noun Phrases are produced with relatively short latencies (probe positions 3 and 8), and though not immediately relateable to the proposed tree structure, this finding is in line with what might be expected if structure is a relevant variable.

#### B. Wilkes and Kennedy (1969)

In this paper, Wilkes and Kennedy set out to develop the notion that major constituents can be separately accessed in a probe task. The paper reports an experiment which is basically the same as Kennedy and Wilkes 1968, but which uses a variety of sentence types in place of a single one. Three sentence types were used, corresponding to the patterns shown in Figure 6-6. Two sentences of each type were tested. Each subject was given a sentence on a card and asked to repeat it aloud and then to recall it without looking at the card. After three perfect repetitions, probes were presented as in the previous experiment. The results of the probe task are shown in Figure 6-7.

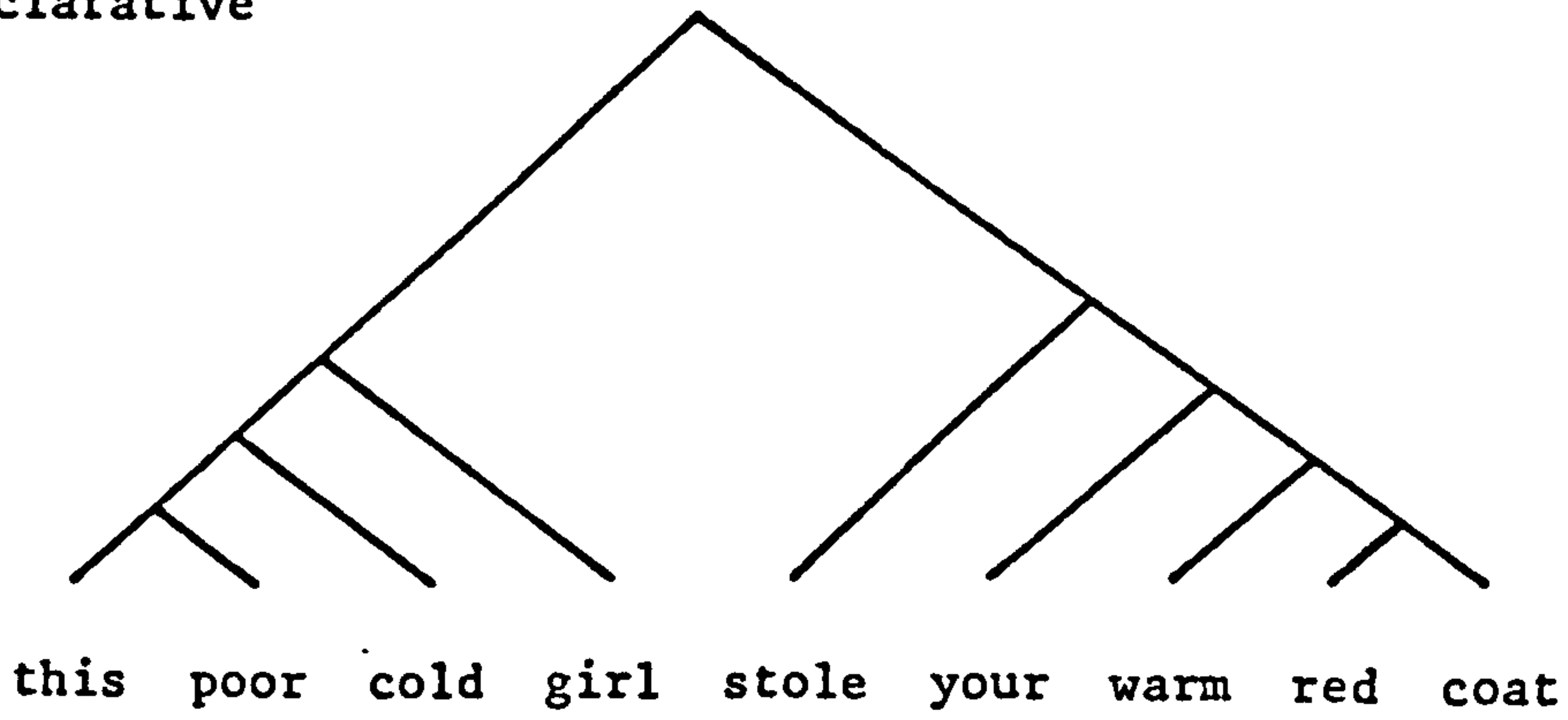
Wilkes and Kennedy report that in all three cases there is a significant effect due to Probe Position. They point out that

Figure 6-6.

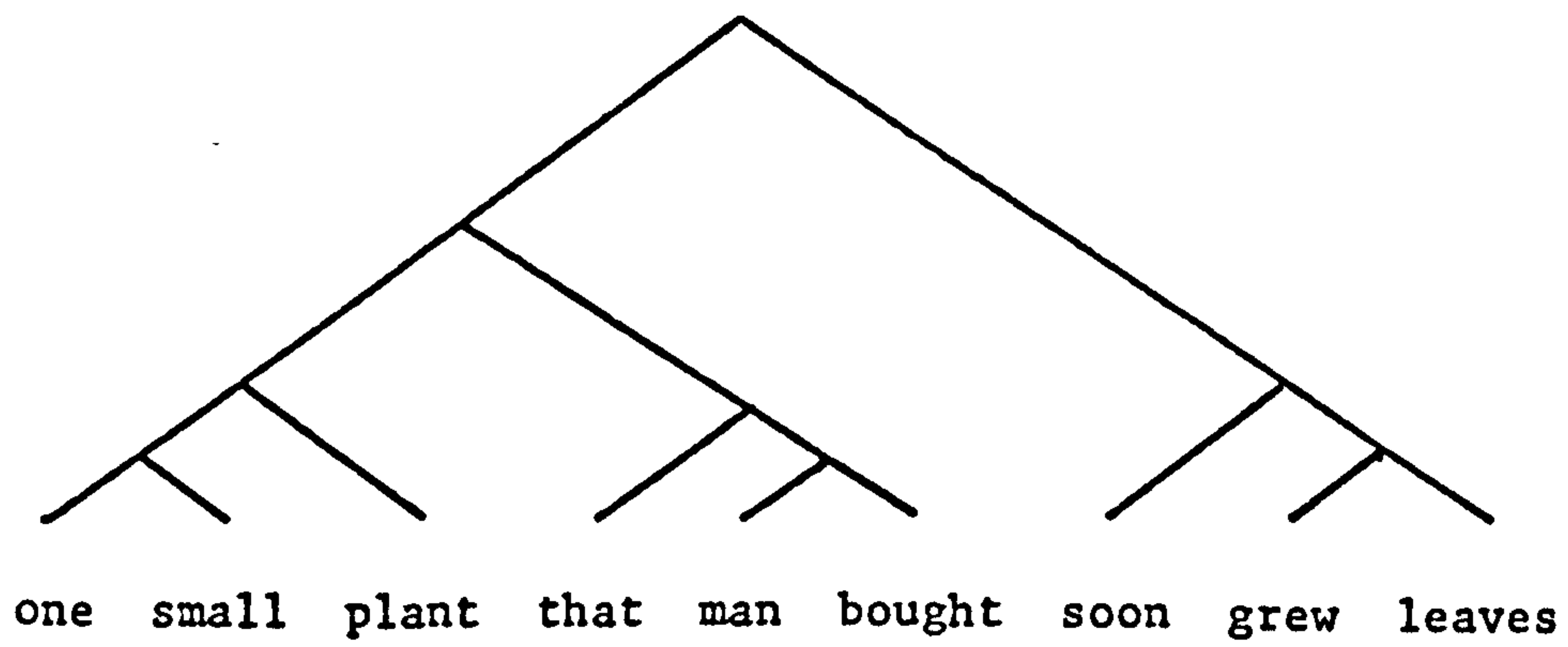
Sentence types used by Wilkes and Kennedy (1969)

The constituent structure is that proposed by Wilkes and Kennedy.

A) Declarative



B) Embedded



C) Multiple

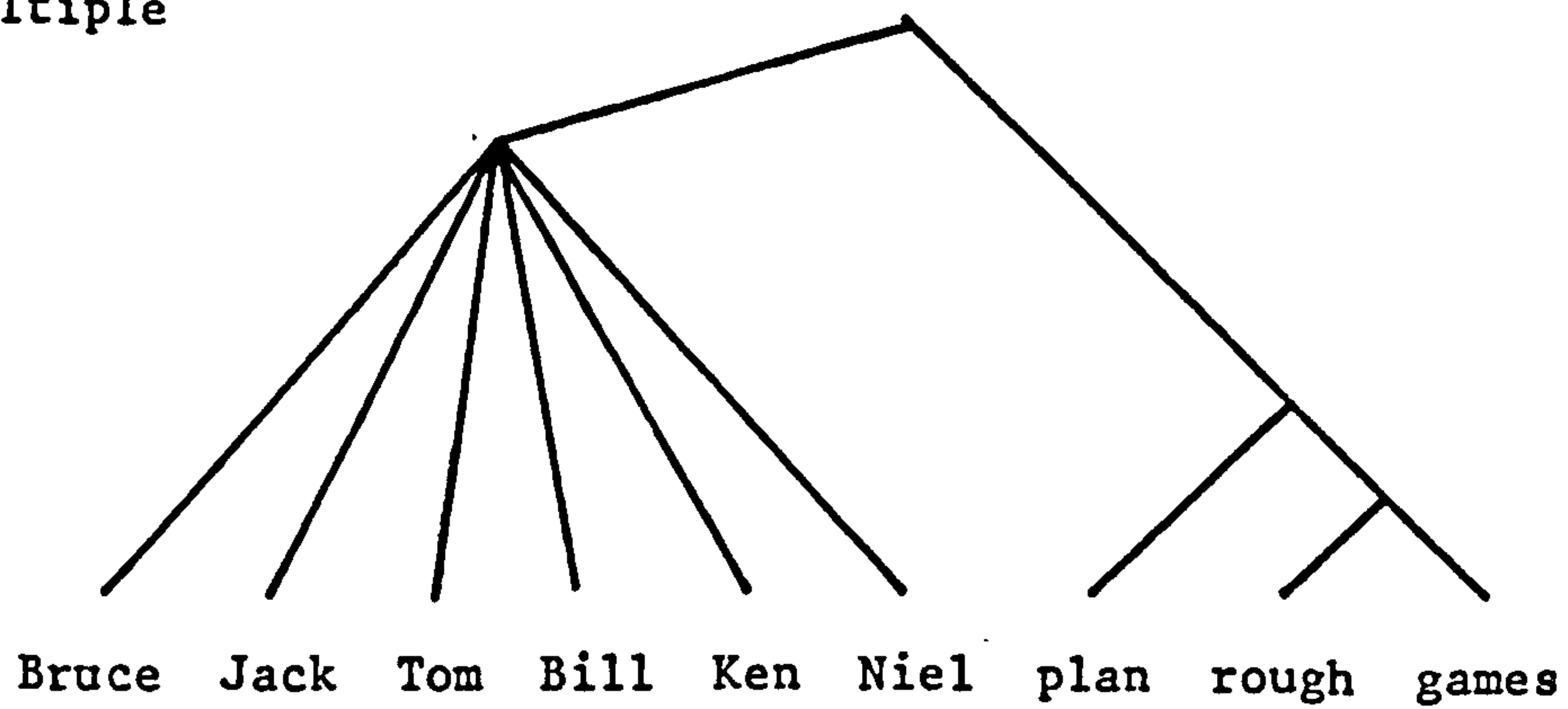
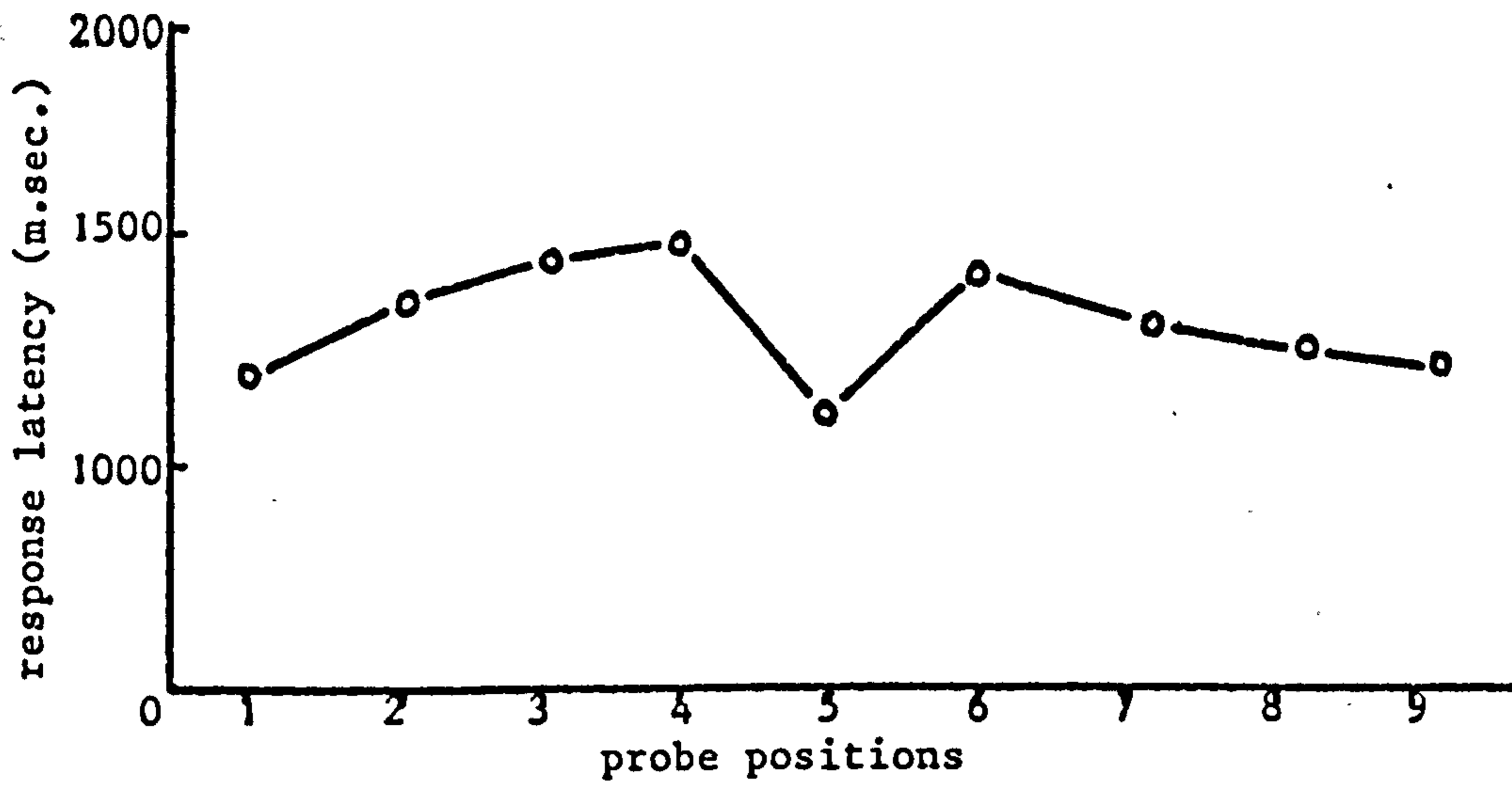




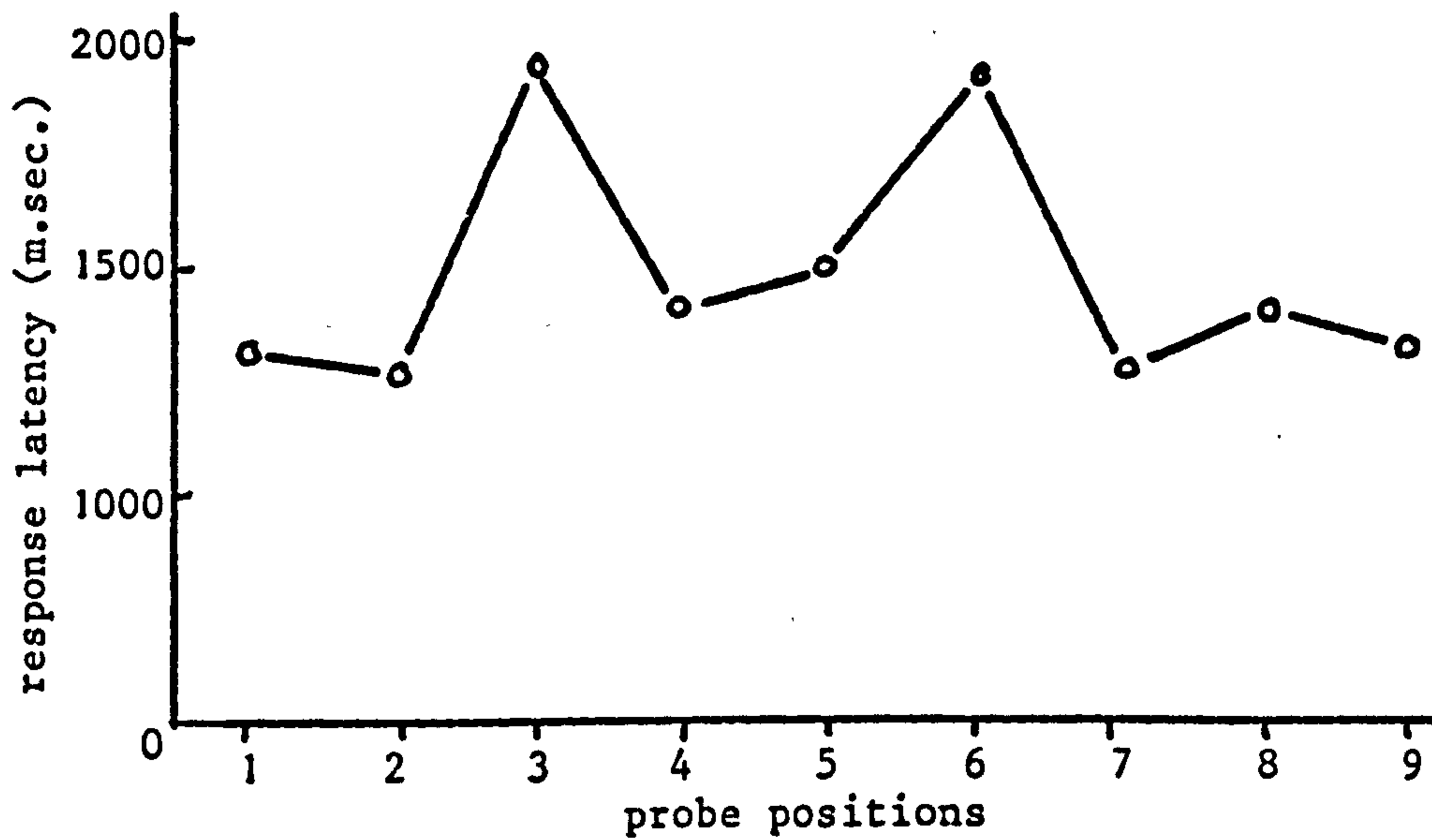
Figure 6-7

Mean response latencies to different sentence types (Wilkes and Kennedy (1969)).

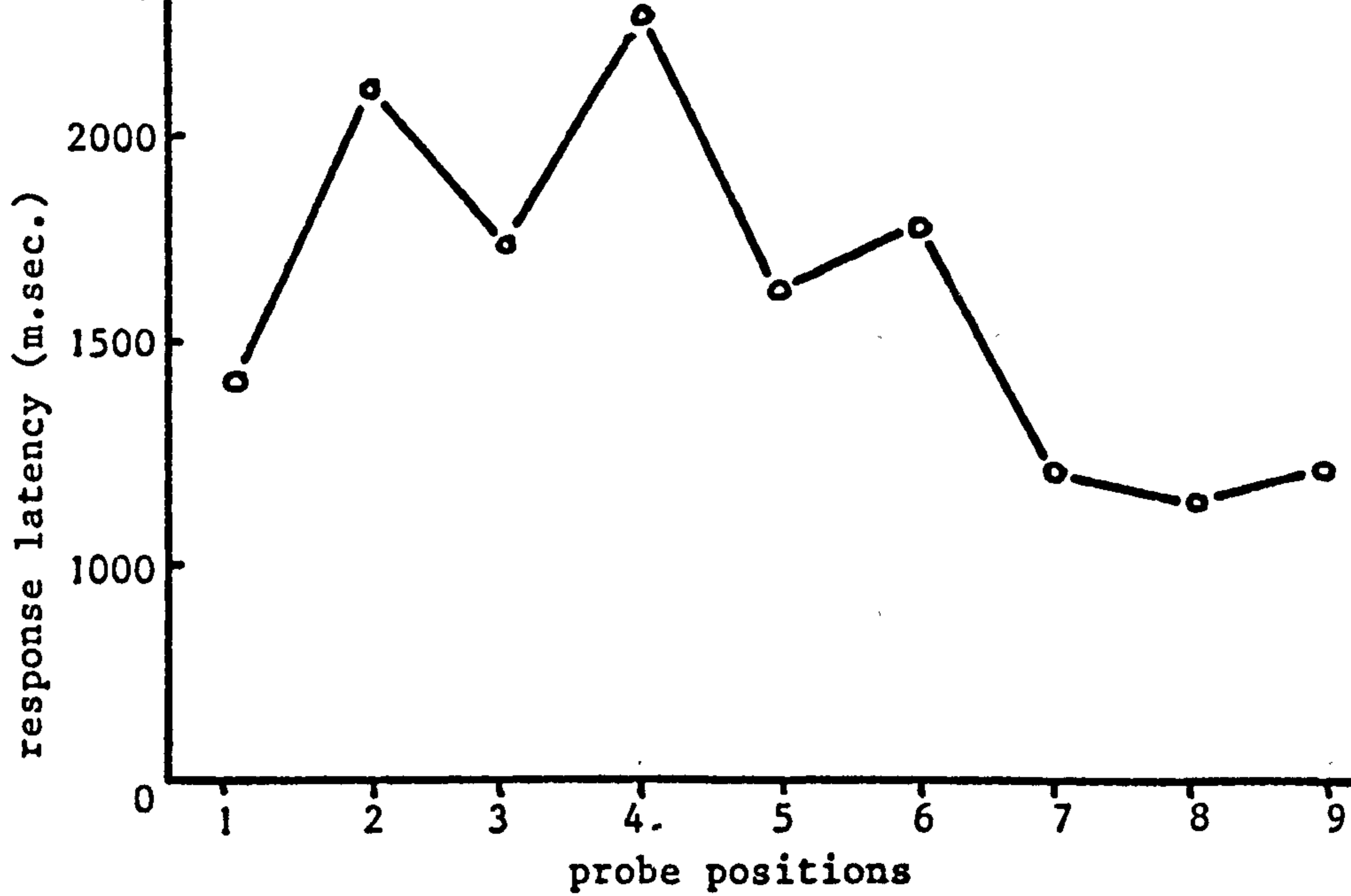
A) Declarative sentences



B) Embedded sentences



C) Multiple sentences



"relatively short latencies tend to fall at the first words of major constituents." The usual major constituent boundary effect is found with the declarative and embedded sentences, but not with the multiple sentences. The last finding is not surprising, however, as this sentence type is exceedingly artificial, and the proper constituent analysis of lists of Noun Phrases of the type included in this sentence is not clear. Wilkes and Kennedy also report that there is a close correspondence between pause patterns while reading and latency, in as much as extreme retrieval latencies, either high or low, appear to be associated with positions where there is marked pausing. Probe words which precede long pauses tend to produce long retrieval latencies, while probe words following long pauses appear to produce relatively short retrieval latencies.

#### C. Kennedy and Wilkes - later work.

The subsequent work of Kennedy and Wilkes (e.g. Kennedy and Wilkes (1969) and Wilkes and Kennedy (1970)) moves away from the central preoccupations of Suci, Ammon and Gamlin's work, and thus does not add significantly to the ideas discussed above. Kennedy and Wilkes develop the idea of entry points, now formally defined as "that position in a sentence which elicits a significantly faster response than the position immediately preceding" (Wilkes and Kennedy (1970)) and pursue this notion rather than the idea of a direct relationship between latencies and syntactic structure as put forward by Suci, Ammon and Gamlin. This leads them to focus their attention on particularly short latencies rather than the long

latencies that are the principal object of Suci, Ammon and Gamlin's consideration, and this means that it is not always easy to extrapolate from Wilkes and Kennedy's data, or to assess the implications of their subsequent experiments. Kennedy and Wilkes do in fact explicitly deny that there is any connection between their entry points and syntactic structure, arguing that these entry points do not appear to be related to either the surface constituent structure or any deep structure relationships in the sentences examined, but this lack of a correspondence may be at least partially due to the fact that entry points are not defined in absolute terms, but only in terms of a difference between adjacent probe positions. The absolute lengths of the latencies are irrelevant to this, of course, and since there is nothing in Suci, Ammon and Gamlin's data to suggest that the first derivative of the latency curves should be related to syntactic structure in any way, this evidence does not disprove Suci, Ammon and Gamlin's claim. Kennedy and Wilkes further report that, though their entry points are not wholly predictable from structure, Auxiliary Verbs and the first Adjective of Object Noun Phrases seem to function as rapid entry points in a wide range of sentence types, and this claim seems to support the idea that some structural factors do influence latencies.

### 6-3-3 Ammon (1968)

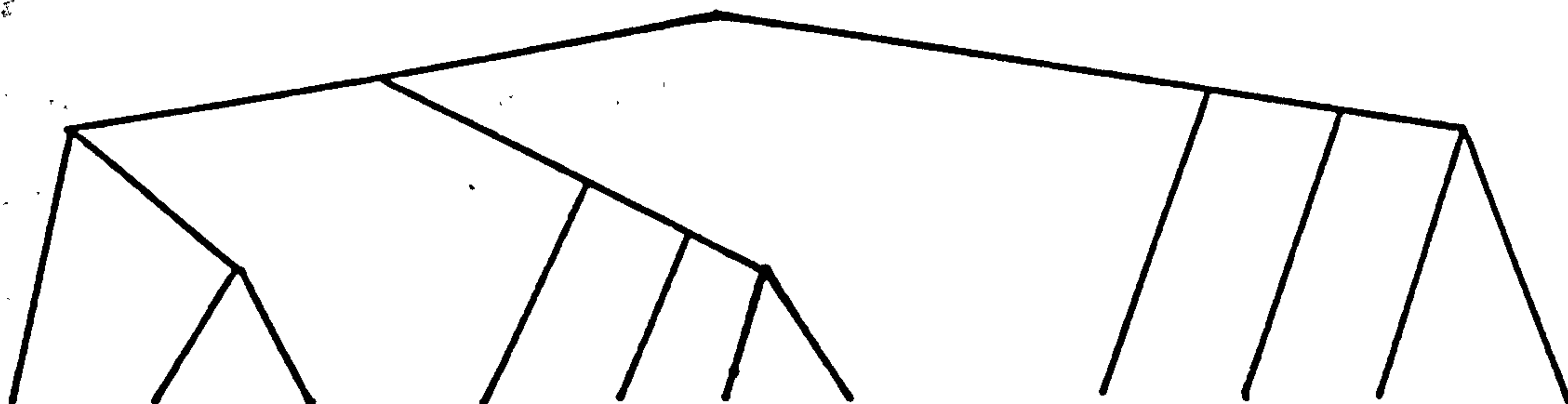
Ammon's study is a replication of the original study by Suci, Ammon and Gamlin (1967), but it makes use of a larger number of sentence types, and uses sentences which are much longer than those of Suci, Ammon and Gamlin. Ammon investigated three different structures, which are shown in Figure 6-8. The constituent structures shown in this figure are those proposed by Ammon, an important point, since Ammon's main purpose in this experiment is to test the idea that latency is predictable from the number of nodes separating a response from its probe word. Two groups were tested: adults and nine year old children. The method follows that of Suci, Ammon and Gamlin, except that Ammon measured response latency from the end of the probe word to the beginning of the response. The other authors discussed so far all used the onset of the probe to the onset of the response as the basic measure of latency. This point is further discussed in Section 6-5. Despite using longer sentences, only five probe positions were studied. These are shown by asterisks in Figure 6-8.

Ammon's results are rather difficult to interpret. There are two main reasons for this. Firstly some of the tree diagrams that Ammon provides are dubious. This applies particularly to types I and II, where only by orthographic criteria could the string be considered as a single sentence. Most people would consider it more natural to treat the two clauses as separate sentences. The other sentence types are technically acceptable, but Ammon fails to point out that there is considerable disagreement among linguists as to

Figure 6-8.

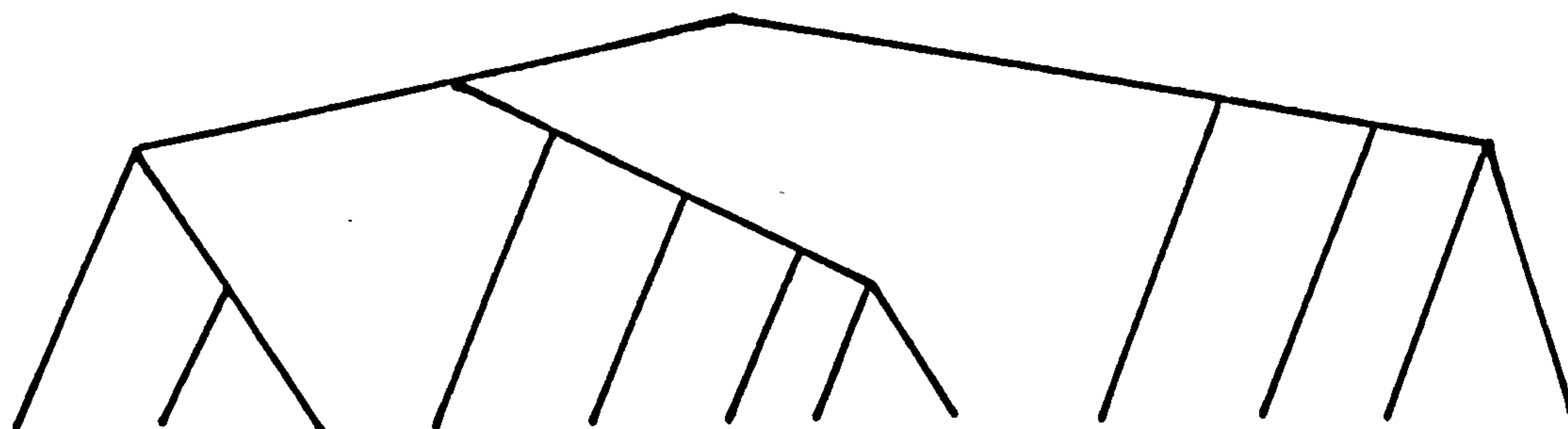
Sentences used by Ammon (1968). Probe positions shown by asterisks. The constituent structure is that proposed by Ammon.

Types I and II



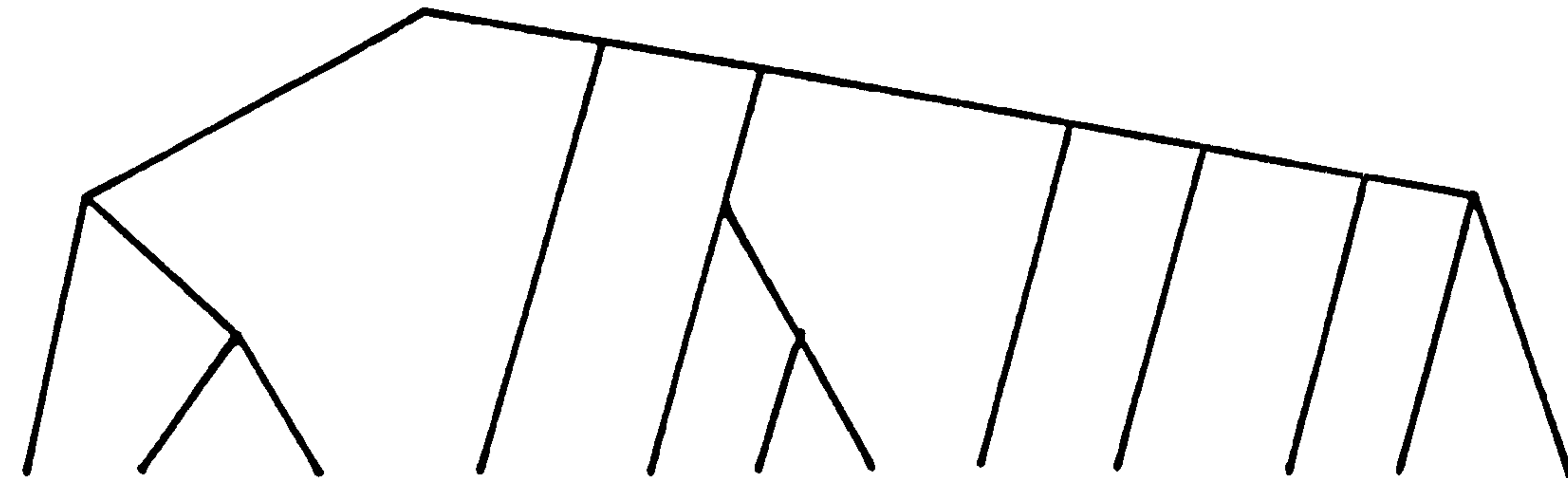
the polite\*actor\*thanked the old\*woman:\*he\*carried the black umbrella

Types III and V



the polite\*actor\*who\*thanked the old\*woman\*carried the black umbrella

Type IV



the polite\*actor\*thanked the old\*woman\*who\*carried the black umbrella

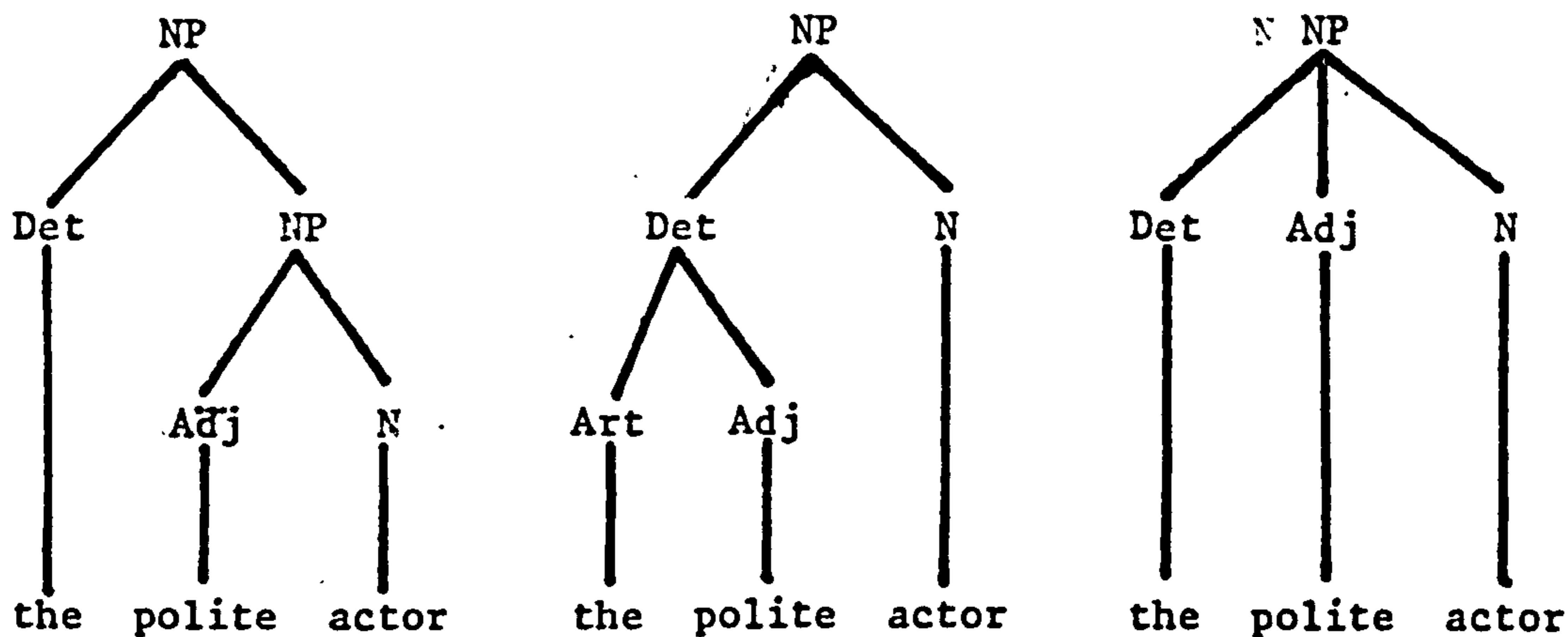
how some of the constituents should be handled. A case could be made, for instance, for any of the three analyses of the polite actor shown in Figure 6-9, but each analysis produces a different measure of the number of nodes between polite and actor, and it is not clear why Ammon's analysis is the one chosen, and no justification of this apparently arbitrary choice is made. Secondly, the scores that Ammon provides for the purposes of discussion are not raw scores, but rather scores that have been transformed in order to correct for lack of homogeneity in the data. Interpretation of the figures is made more complicated by the fact that the transformation has the effect of inverting the scores, long raw latencies come out as short transformed latencies, while short raw latencies are converted into large transformed latencies. It is not at all apparent why Ammon chose this particular transformation formula ( $1000 / (\text{raw score} + 5)$ ) to solve his homogeneity problem. Its main effect is to bring together scores above 1.12 seconds, while expanding scores below this point. The figure of 1.12 seconds seems to be the shortest mean latency produced by the children; the adult means are normally shorter than this figure. It appears, then, that Ammon's formula will increase the spread of the adult scores, while reducing that of the children's scores. The implications of this for evaluating the effect are unclear, since one of the effects of non-linear transformations like this one, is to change the interpretation of interactions. (cf. Smith (1976)). This problem is not discussed by Ammon.

Ammon reports the following findings, which in view of the discussion above, should perhaps be treated with some caution. All the sentence

types produced a significant difference between the two groups, in that the adults produced shorter raw latencies. All sentence types produced a significant effect for probe position. In addition, Type III sentences produced a significant Age x Probe Position interaction, but this appears to be due to small differences which do not affect the overall pattern. In all cases, the longest latency is produced by probe-response pairs that cross major constituent boundaries. In addition, if the Transformed Latency scores are converted back into raw scores, the results show differences between adults and children which are very similar to those reported by Suci, Ammon and Gamlin. The children seem to produce patterns very much like those of the adults, except that the difference between the probe positions in the children's data are more exaggerated.

Figure 6-9.

Three possible phrase structure analyses of 'the polite actor'



6-3-4 Kempen (1976).

Like Kennedy and Wilkes, Kempen also makes a number of changes in the classical method. Kempen asks his subjects to remember four sentences at a time, each of which share a common syntactic structure. Examples are shown in Figure 6-10 below. Probe words from all four sentences are then presented in a random sequence, until all the main probe positions of all the sentences have been tested. Kempen provides a full statistical treatment for only two probe positions: the between constituent position and the within predicate position. These are shown by asterisks in Figure 6-10. Kempen's material is in Dutch, but the syntactic patterns are broadly similar to those of the original Suci, Ammon and Gamlin experiment. The results of this study, also shown in Figure 6-10 show the familiar slower latencies at the between constituent position. The difference is about 150 m.secs. Kempen also reports three further experiments, but these do not significantly add to the discussion.



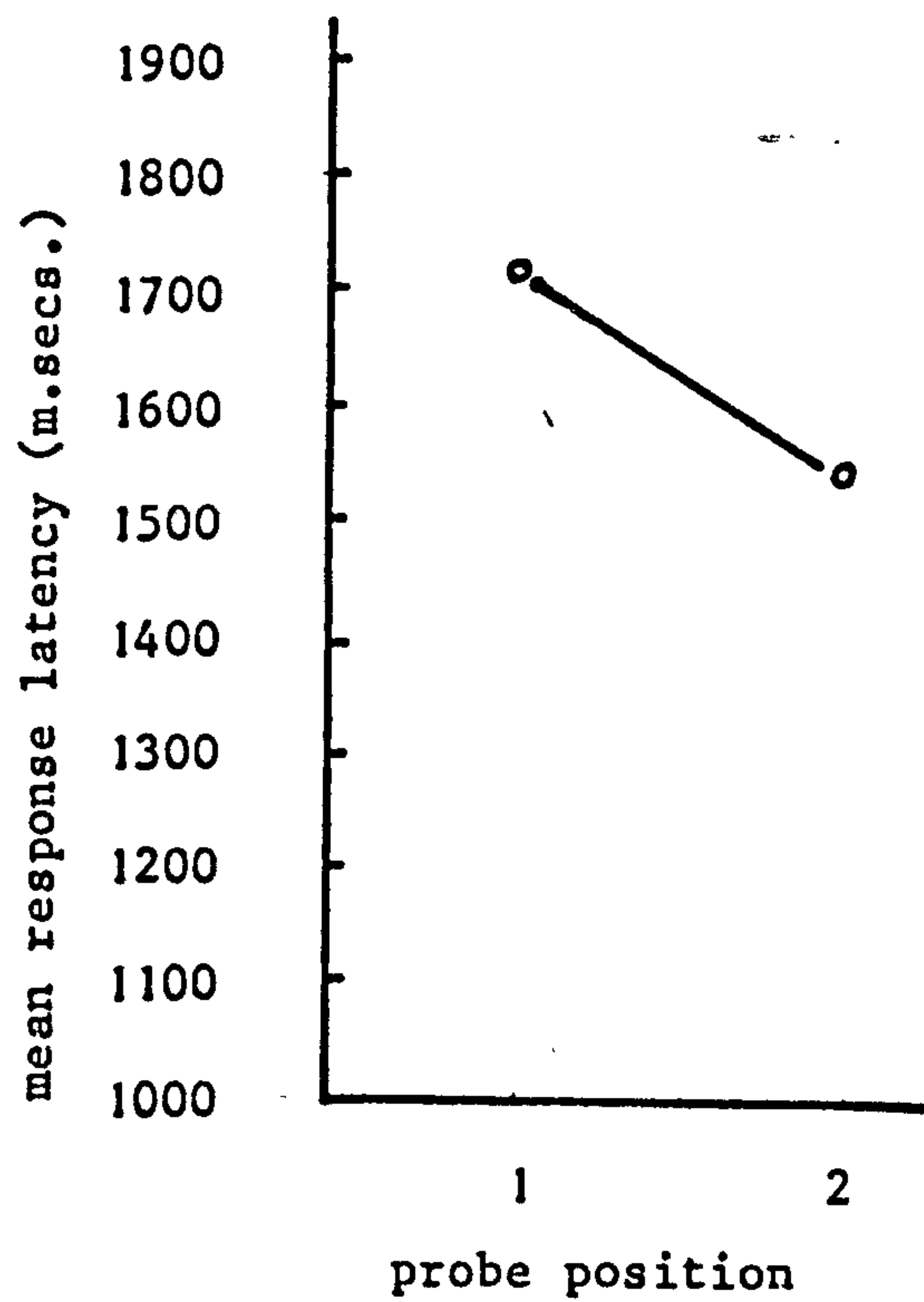
Figure 6-10.

A) sentences used by Kempen (1976)

Different sentences were given to each subject, but all were of the same syntactic type, and differ only in the lexical items used. An example set is given below.

STRUCTURE (Article (Adj (Noun))) \* (Verb \* (Noun)) (Adverb)  
SAMPLES Die twee Belgen lazen boeken keurig  
Die drie Finnen leerden teksten moeizaam  
Die vier Grieken schreven zinnen matig  
Die vijf Ieren typten worden prettig

B) Mean response latency to different probe positions (Kempen (1976))



#### 6-4. PROBE LATENCIES AND SECOND LANGUAGE SPEAKERS.

No probe latency studies exist which have used non-native speakers as subjects. The relevance of the technique to the question of how non-native speakers of a language react to the syntactic structure of its sentences is not difficult to see, however. If the learners are generally insensitive to the higher levels of syntactic organization in foreign language sentences, then they ought to handle all sentences in an identical fashion, irrespective of their syntactic structure. This argument would lead one to predict that a pair of sentences such as those shown in Figure 6-1, which produce strikingly different latency curves in native speakers would fail to do so in a group of non-native speakers of English. This latter group should produce two identical curves. Though at this stage it is not possible to predict what the shape of these curves should be, one might hazard a guess that the curves might be related to serial position, as there is some indication that probe latencies measured for lists of unrelated words vary according to the serial position of the probe word. (Kennedy (1968)).

On the other hand, if the Contrasting Structure proposal is sound, then a rather different type of response pattern might be expected to emerge. If non-native speakers can handle structures which occur in their own language, then they ought to produce latency curves which are broadly similar to those produced by

native speaker subjects in that they should clearly discriminate between sentence types with differing structures. At the same time, however, some major deviations from the native speaker patterns might be expected at points of contrast between the two languages. In the case of Figure 6-1 for instance, a group of native Spanish speakers, say, would be expected to satisfactorily discriminate between the two sentence types, but to have abnormally long latencies in probe positions 1 and 4 in Sentence I and in probe position 2 in Sentence II, positions where Adjective probes are followed by Noun responses. Since preposed Adjectives of this type contrast with the predominantly post-posed adjectives preferred by Spanish, these positions might be expected to be a source of difficulty for the Spanish speakers. All the other structure types do occur in Spanish, and these should not therefore be problematical in any serious way.

On the face of it, then, the Probe Latency Technique offers a straightforward way of testing the contradictory predictions made by the Revised Word-List Conjecture and the weaker Contrasting Structure Conjecture. A series of Experiments designed to assess these predictions is reported in the following sections.

6-5.        EXPERIMENT 1.

This study was an exploratory one undertaken in order to verify that Probe Latency effects of the sort described by Suci, Ammon and Gamlin (1967) for English are also to be found with native speakers of Spanish in their own language. There is no reason to assume that this would not be the case, but since it was intended to use Spanish as the main language of testing, it seemed important to establish that this was the case. There would, clearly, be little point in arguing that learners should distinguish between different structures in the Probe Latency test if native speakers of the language systematically failed to do so.

## 6-5-1. Materials.

The materials tested in this first study consisted of fifty Spanish sentences. Two syntactic structures were investigated, and details of these will be found in Table 6-1. These two structures accounted for 16 sentences each. The remaining 18 sentences were of various syntactic types, and were included to provide practice sentences and filler sentences designed to prevent Subjects from developing a set for a small range of syntactic structures. A complete list of the stimuli tested will be found in Appendix 6-A-1. The fifty sentences were recorded on tape by a woman with a Castillian Spanish accent. Each sentence was followed by a probe word, so that every sentence type was probed four times in each of the four probe positions.

The conjectures we are studying lead us to make a number of fairly straightforward predictions about the behaviour of native speakers and learners on a probe latency task using materials of this type.

In the first place, we would expect the native speakers to distinguish between the two sentence types, and they should produce longest latencies in probe position 2 for the A sentences, and probe position 1 for the B sentences, since these positions correspond to the major constituent boundaries of the two sentence types. In the second place, since the two sentence types do not contain any instances of short, high frequency grammatical function words being used as probe words, the Revised Word-List Conjecture predicts that a group of non-native speakers should fail to distinguish the two sentence types, producing essentially similar curves in both instances.

Table 6-1.

Sentence Types probed in Experiment 1  
Probe Positions are shown by asterisks.

TYPE A

STRUCTURE: (Article Noun \* Adjective) \* (Verb \* (Noun \* Adjective))

EXAMPLE: La muchacha tonta comió manzanas verdes

LITERALLY: The girl stupid ate apples green

I.E: the stupid girl ate green apples

TYPE B:

STRUCTURE: (Article Noun) \* (Verb \* (Adjective \* Noun \* Adjective))

EXAMPLE: La secretaria escribió cuatro cartas importantes

LITERALLY: The secretary wrote four letters important

I.E. The secretary wrote four important letters

In contrast to this, if native English speakers are tested, the Contrasting Structures Conjecture would lead us to expect very similar curves to those produced by the native Spanish speakers, but with possible discrepancies and minor deviations from the native speaker curves. These discrepancies would be expected in probe positions 1 and 4 for the A sentences, and in probe position 4 for the B sentences. All these positions have Noun probes with Adjective responses. As we have seen, this use of postposed adjectives is one of the major points of contrast between English and Spanish, and might thus be expected to produce exceptionally long latencies. In addition, we might also expect to find exceptionally long latencies for probe position 2 in the A sentences, where a postposed adjective serves as a probe for a following Verb. This particular transition can only occur in languages that have postposed adjectives, and so is extremely rare in English. However, since this probe position coincides with the major constituent boundary of the sentence, where long latencies would be anyway expected, this effect might be less noticeable than if the contrast had occurred in a different position.

### 6-5-2 Subjects.

Twelve native speakers of Spanish were tested. These subjects were all following courses in London University of in Language Schools in Central London. A small group of students (N=5) studying Spanish in their first year of a degree course at the Polytechnic of North London was also tested as a preliminary control.

### 6-5-3 Method.

Each subject was tested individually in a single session that lasted approximately 25 minutes. Subjects heard each sentence spoken from the tape, and after a pause of one second, the probe word was projected onto a screen placed four feet in front of the subject. Response latencies were recorded by a digital timer that was activated by the presentation of the slide, and stopped by the subject calling his response. This method of presentation is slightly different from that used by Suci et al (1967). Their experiment used auditory presentation of both sentence and stimulus probe, and took the onset of the probe to the onset of the response as their measure of latency. This measure seems to be rather unsatisfactory, since it takes no account of the variations in the length of the probe words. Suci, Ammon and Gamlin's probes are all monosyllabic, but even within this constraint, there is considerable variation in word length, systematic differences occurring, for example, between syllables with long or short vowels or diphthongs, between syllables containing final voiced or unvoiced consonants, or final stops or fricatives, and

between syllables that contain final consonant clusters or not. It is not clear that Suci, Ammon and Gamlin are entirely justified in ignoring this problem, since it produces variation in the latencies which is not random. A second problem with the auditory method of presentation is the danger of setting up an underlying rhythm by which the response latency might be affected. This danger is particularly serious when the test sentence consists of monosyllabic words, and a clear intra-sentence rhythm is established. This is a factor which subjective reports from subjects who have experienced auditory presentation of material indicate may be a factor of some importance. It appears to have received little attention from those working with response latencies of any kind to material presented in the auditory mode, and is not normally considered to be a factor that needs to be controlled when materials are being prepared. A third problem is that there is always some degree of inaccuracy in using detailed measurement of reaction times as a dependent variable when these are based on speech stimuli. This is because speech sounds do not always have a clear onset, and some low amplitude sounds such as nasals, often fail to trigger a voice key even if the triggering level is not very low. (Low settings, of course, run the additional risk of allowing the timer to respond to incidental sounds such as breathing, and this can seriously reduce the reliability of the data. Some uncertainty of this kind is clearly inevitable in the probe latency technique, but using an auditory presentation of the probe words seems to compound the problem by introducing uncertainties in both the reference points of the latency scores, and thus multiplying the possible margin of error by a factor of two.



Visual presentation of the probe words was adopted for this experiment, as it seemed to offer a way of avoiding most of these problems. Visual presentation makes it very easy to determine accurately the onset of the probe, in a way that is not dependent on any of the physical characteristics of the probe word, such as its first sound. Visual presentation avoids the setting up of auditory rhythms. Furthermore, visual presentation makes the length of the stimulus word less important, and this is a major consideration in an experiment that uses Spanish material, since few words in the language are monosyllabic, and probe words must of necessity be longer than the probe words used in Suci's original experiment. This last consideration effectively rules out the possibility of using auditory probes, since the time necessary for pronouncing the probe word would account for a major part of the response latency if two or three syllable words were used as probes, and this would make the response latency measure proportionately unreliable.

Visual presentation of the Probe word had been used previously by Kennedy and Wilkes (1968) without grossly affecting the overall pattern of results found, and Kornfeld (1972) reported that cross-modal testing of the kind adopted here actually increases the effect of probe position. The use of this type of presentation seems therefore to be fully justified, and in no way problematic.

Table 6-2.

Mean median response latencies<sup>17</sup> to four probe positions in two Spanish sentence types; native speakers and learners of Spanish. Response latencies are given in m.secs.

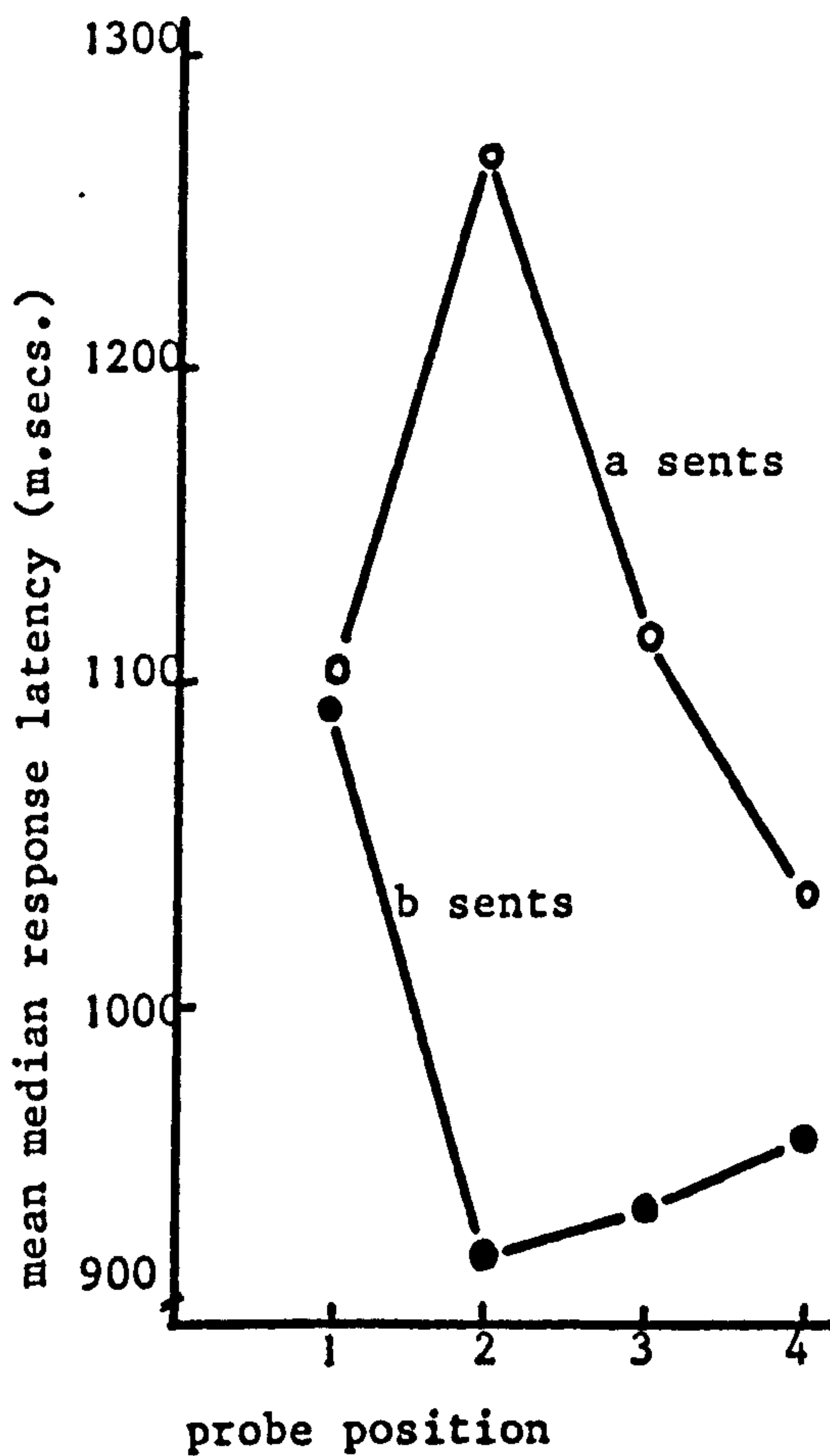
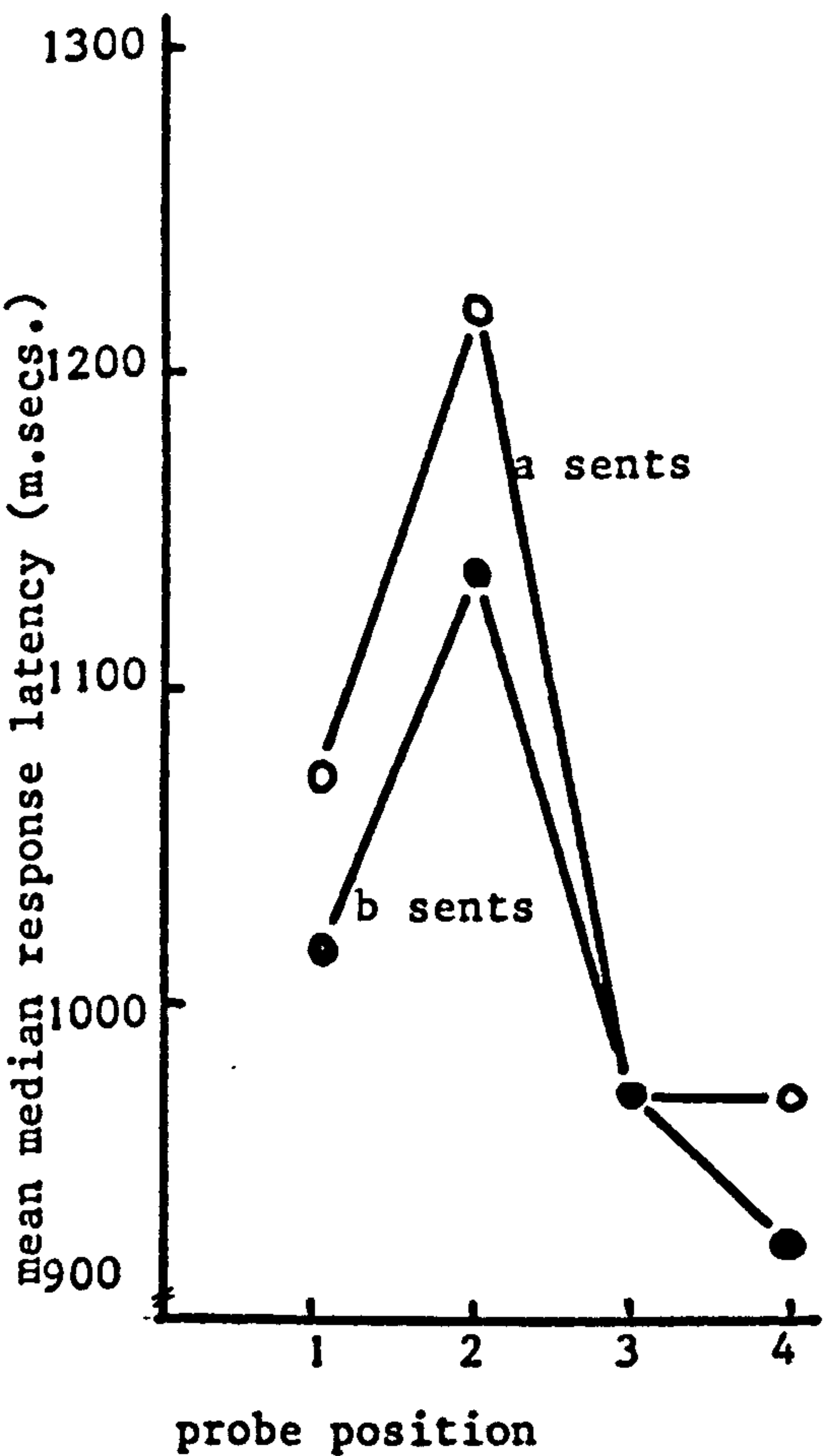
GROUP	Native Spanish Speakers				Learners of Spanish			
	1	2	3	4	1	2	3	4
a sents:	1072	1220	971	971	1107	1267	1118	988
b sents:	1019	1132	973	926	1096	922	932	956

Figure 6-11.

Mean median response latencies for four probe positions in two Spanish sentence types; Native Speakers and Learners.

Native Speakers

Learners



6-5-4.Results.

For each subject, the median of his scores for each probe position of both sentence types was calculated. Errors, non-responses and responses which were insufficiently loud to activate the timing mechanism were not counted in these results. This procedure is the same as that adopted by Suci et al. The full results will be found in Appendix 6A-2. They are shown in summary form in Table 6-2 and Figure 6-11. These results were subjected to two separate analyses of variance, one analysis for each group, in which the main effects were Sentence Type and Probe Position. The results of the analysis are shown in Table 6-3.

Table 6-3.

Summary of Analyses of Variance.

A: Native Speakers

SOURCE	df	SS	MS	F	prob
Subjects	11	5748834			
Sentence Type	1	50508	50508	4.4639	not sig.
Error	11	124464	11314		
Probe Position	3	752898	250996	5.9048	.005
Error	33	1402564	24501		
Sentence x Probe	3	24123	8041	0.427	not sig.
Error	33	621269	18826		
Residual Variance	84	2975829			

B: Learners

SOURCE	df	SS	MS	F	prob
Subjects	4	578071			
Sentence Type	1	206209	206209	3.5996	not sig.
Error	4	229249	57287		
Probe Position	2	113477	37825	0.4679	not sig.
Error	12	502337	41861		
Sentence x Probe	2	181395	60465	4.0634	.05
Error	12	178566			
Residual Variance	35	1411137			

Sums of squares and mean squares are rounded to the nearest whole number.

#### A. Native Speakers

The only significant effect found with the native speaker group is a highly significant effect due to probe position. Both sentence types show similar response curves, and the interaction between sentence types and probe position fails to approach significance.

#### B. Learners.

The results of the learner group show a pattern of responses quite unlike those of the native speakers. Probe position alone is not a significant source of variance, but probe position and sentence type interact to produce an effect significant beyond the .05 level.

#### 6-4-5 Discussion.

These results were totally unexpected. Suci et al's work would have led one to predict that the native speaker group should distinguish the two sentence types by producing latency curves reflecting the differing structures of the two sentence types. In fact, the native speakers' curves are not significantly different, both curves showing a marked peak in probe position 2. This position does correspond to the major constituent boundary of the (a) sentences, but fails within a noun phrase in the (b) sentences.

The learner results were equally surprising. Only five subjects were run in this group, as the experiment was prematurely abandoned when it became apparent that the native speaker results did not conform to the predicted patterns. In spite of this small number, however, the group clearly does distinguish between the sentence types, and furthermore, does so in a manner which bears a close relation to the syntactic structure of the sentences, in that both sentence types produce longest latencies at the major constituent boundary.

These results clearly provide us with no evidence in support of either the Revised Word-List Conjecture or the Contrasting Structures Conjecture that we have set up against it. On the contrary, in fact, the learners' results show clear evidence of a structuring effect which is incompatible with the predictions made by either of these conjectures. The learner results are in fact indistinguishable from what we would have predicted of the native speakers.

However, this result is less surprising than the fact that the native speakers failed to perform in the way that we predicted, and this finding makes it difficult to accept the obvious interpretation of the learners' results without further consideration. Some explanation of these anomalous findings is clearly necessary at this stage.

There are two main types of explanation which might reasonably account for the performance of the native Spanish speakers. The more drastic of these is to query the robustness of the Probe-Latency Technique, and to argue that it may not apply outside the framework in which it was first used, and in particular that it may not hold for Spanish, or for the particular group of Subjects tested here. This is obviously a last resort, however, and so will not be followed up at this stage, although some explanations along these lines will be further discussed in Section 6-7. An alternative, less radical, explanation might be sought in the actual materials used, and, given that the native Spanish speakers did perform as expected on the (A) sentences, where the Major Constituent Boundary produces the longest latencies, a reasonably parsimonious explanation of their failure might be that there are some peculiar features in the (B) sentences which produce anomalous reactions. The materials do provide some grounds for this surmise, in that the final Noun Phrase construction ( $_{NP}$ (Adjective Noun, Adjective)) - is only possible in Spanish if one of a relatively restricted set of Adjectives appears in the initial position. This set includes numerals, quantity words such as mucho, poco, and varios, ('many', 'few', 'several'). It is these words that

give rise to the unexpectedly long latencies in this curve, when they appear as responses to the probe word in second position. Possibly the fact that a transition from a Verb to an Adjective is relatively rare in Spanish may have caused the long latency here. Alternatively, perhaps the essentially arbitrary link between these words and the Nouns which follow them may have caused interference from one sentence to another, and thus caused long latencies in this position as an artefact. A third explanation is that some of these words were used more than once (although none of them was used as a response or as a probe more than once) and this too may have caused interference from one sentence to another. These explanations are not wholly satisfactory. They are essentially ad hoc, but more importantly they fail to account for the results produced by the Learner Group. If considerations such as repetition affect the performance of native speakers, one might expect them to have an even greater effect on the performance of non-native speakers, who are generally rather more sensitive to the actual wording used than native speakers are. Nevertheless, the suggestion is sufficiently plausible for it to appear to be worthwhile exploring before more far-reaching possibilities were considered, and for this reason a second experiment which attempted to rule out these explanations by using a different range of materials was carried out. This experiment is reported in the next section.

6-5-1 Introduction

This experiment was a replication of the experiment reported in the previous section. New materials were constructed which did not contain the problematical long Noun Phrases used in that experiment.

6-5-2 Materials.

Forty Spanish sentences were constructed. Sixteen of these sentences were miscellaneous, of a variety of syntactic types. These sentences were used as filler items designed to prevent Subjects from developing a set for particular sentence types. The remaining 24 sentences were of two types as shown in Table 6-4. A complete list of the stimuli will be found in Appendix 6-B-1.

Table 6-4.

Sentence Types probed in Experiment 2.  
 Probe Positions are shown by asterisks.

TYPE A:

STRUCTURE: (Determiner Noun) \* (Verb \* (Noun \* Adjective))

EXAMPLE:                    la   mújer    llevaba zapatos negros  
 LITERALLY:                the woman was wearing shoes black  
 I.E.                        the woman was wearing black shoes

TYPE B:

STRUCTURE: (Determiner Noun \* Adjective) \* (Verb \* Noun)

EXAMPLE:                    el   negociante rico    bebía coñac  
 LITERALLY:                the business-man rich was drinking brandy  
 I.E.                        the rich business-man was drinking brandy



These sentences differ from the previous experiment in that the long final Noun Phrases are avoided, and these sentences contain five words and three probe positions. The only major structural difference between the sentence types lies in the position of the adjective - after the Object Noun in the Type A sentences, after the Subject Noun in the Type B sentences.

These sentences were recorded on tape by a woman with a Castillian accent. Each sentence was followed by a probe word, presented visually, so that both sentence types were probed four times in each of the three probe positions. Probe words were selected so that correct responses began with high amplitude sounds which stopped the timing mechanism with maximum efficiency.

#### 6-6-3 Subjects.

Two groups of subjects were tested: a group of eleven native speakers of Spanish, all resident in Britain, and following courses in London University or London Language Schools; and a group of sixteen native English speakers studying Spanish at A-level or first year degree standard.

#### 6-5-4 Method.

The method followed was the same as in the previous experiment. Each of the subjects was tested individually in a session lasting approximately 25 minutes. The Subject heard each sentence spoken on tape, and after a one second pause, the probe word was flashed onto a screen in front of him. The Subject's task was to say as quickly as possible the word in the sentence that followed the probe word. The dependent variable measured is the latency of this response, which is taken as the time elapsing between the presentation of the probe word to the onset of the response.

Once again, the conjectures we are investigating make relatively straightforward predictions about the performance of the learner group and how it should differ from that of the native speakers. The Revised Word-List Conjecture predicts that the learners should treat both sentence types in an identical fashion, and fail to discriminate between them. The Contrasting Structure Conjecture suggests that the non-natives should distinguish between the sentences in much the same way as the native speakers by producing their longest latencies in the probe positions predicted by the syntactic structure of the sentences. This Conjecture also predicts that exceptionally long latencies may occur in probe position 3 for the A sentences and in probe position 1 for the B sentences. These positions all consist of Noun probes followed by Adjective Responses. The native speakers, of course, are expected to produce their longest latencies at probe position 1 in the A sentences, and in probe position 2 for the B sentences, since these positions correspond to the Major Constituent Boundaries of the test sentences.

Table 6-5.

Mean median response latencies to three probe positions in two Spanish sentence types.

PROBE POSITION	1	2	3
Native Speakers	1024	1073	992
Type A	1024	1073	992
Type B	1018	1042	975
Learner Group			
Type A	1188	1152	1046
Type B	1101	1348	1149

Figure 6-12.

Mean median response latencies to three probe positions in two Spanish sentence types.

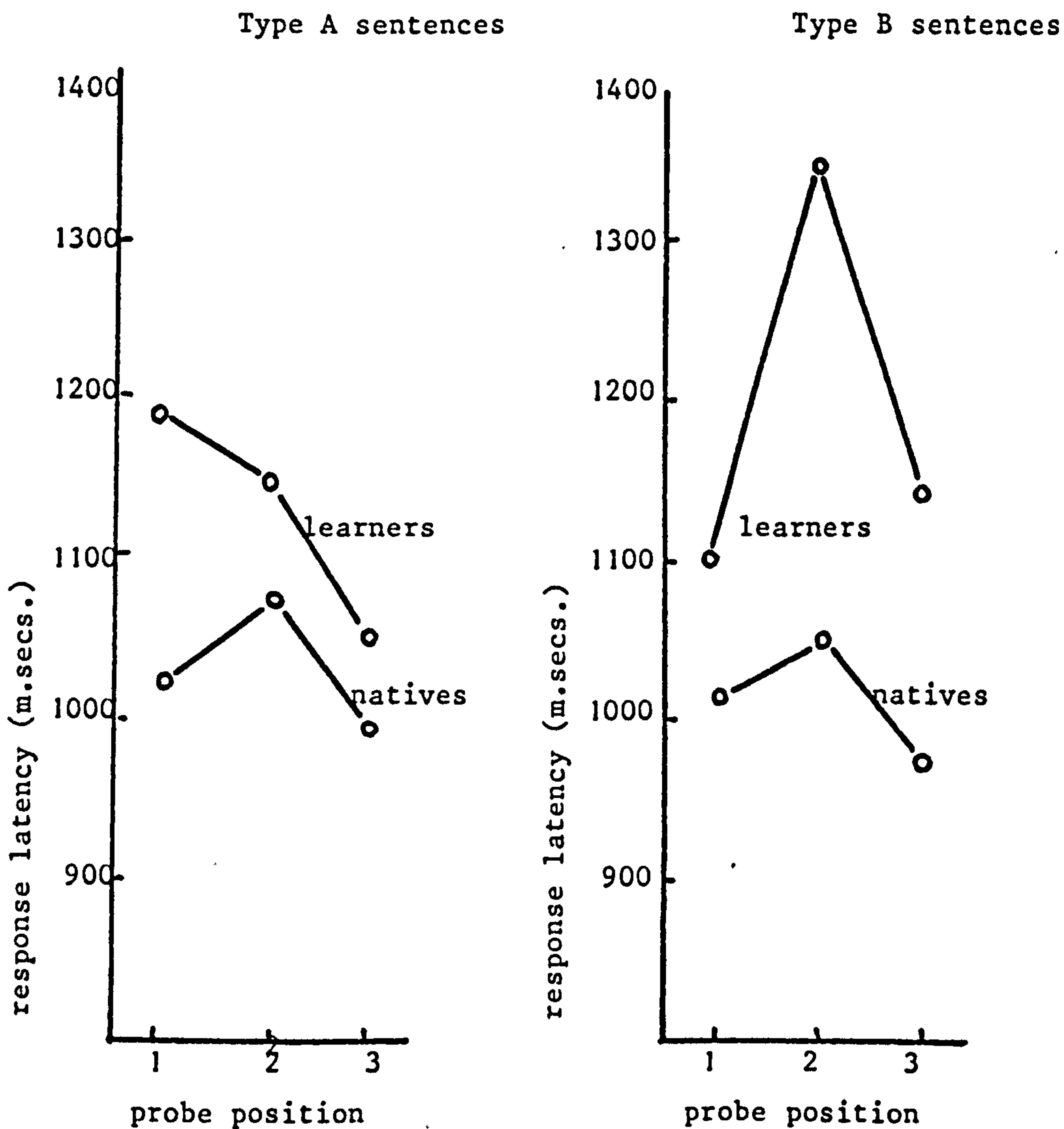


Table 6-6.

Analyses of Variance: Experiment 2.

A: Learners

SOURCE	df	SS	MS	F	prob
Subjects	15	1816418			
Sentence Type	1	89121	89121	2.4028	.1419
Error	15	556358	37090		
Probe Position	2	313034	156517	4.0203	.0284
Error	30	1167940	38931		
Sentence x Probe	2	277964	138982	4.2459	.0237
Error	30	981994	32733		
Residual Variance	80	3386412			

B: Native Speakers

SOURCE	df	SS	MS	F	prob
Subjects	10	2007066			
Sentence Type	1	5364	5364	0.7166	.417
Error	10	74848	7485		
Probe Position	2	58914	29457	1.2748	.301
Error	20	462160	23108		
Sentence x Probe	2	1619	809	0.1000	.905
Error	20	161856	8093		
Residual Variance	55	764762			

Note: Some of the squares and mean squares are rounded to the nearest whole digit

## 6-6-5 Results.

For each subject, the median of his scores for each probe position was calculated. Errors, non-responses and responses which were insufficiently loud to activate the timing mechanism were not included in these figures, and a discussion of the error patterns will be found in Appendix 6-B-3. The full results will be found in Appendix 6-B-2, and they are shown in summary form in Table 6-5 and Figure 6-12. The results were submitted to two separate analyses of variance, one for each group of subjects. The main effects in these analyses were Sentence Type and Probe Position. Summaries of the Analyses of Variance will be found in Table 6-6.

These results are essentially the same as those reported in Section 6-5. The main difference between the groups is that the learner group performs at a slower level than the native speaker group. This is probably due to the fact that a number of less advanced learners were used in this study. In other respects, however, the results are comparable. Contrary to prediction, the learner group does distinguish between the sentence types, and their latency scores are closely related to the constituent structure of the sentences, in that the Major Constituent Boundary produces the longest latency in both cases, and Probe-Response pairs lying wholly within a Noun Phrase produce the shortest latencies. The analysis of variance for the learners alone showed the Sentence Type x Probe Position interaction to be significant ( $F=4.2$ ,  $p<.024$  with 2,30 df). In contrast, the native speaker group do not distinguish between the Sentence Types. Rather,

the longest latencies are produced by the second probe position, just as in Experiment 1, irrespective of Sentence Type. For this group, the Interaction between Sentence Type and Probe Position does not approach significance ( $F=0.1$ ,  $p>.05$  with 2,20 df).

## 6-7 DISCUSSION.

### 6-7-1. Some Possible Explanations.

The results of Experiment 2, like those of Experiment 1, fail to produce any evidence in support of the two conjectures under test. In both cases, the learner group have produced data which corresponds exactly with what would have been expected of the native speakers, with longest latencies being produced where probe and response straddle a major constituent boundary. This data clearly reflects the structure of the stimulus strings, and thus argues strongly against the Revised Word-List Conjecture. At the same time, there is no evidence to indicate that structural contrasts between English and Spanish have any adverse effect on the performance of the learner group. In both of the stimulus sentence types, the Noun-Adjective transitions (A3 and B1) are the fastest latencies of all. The Adjective-Verb transition of the B-sentences (B2) does produce an exceptionally long latency, but without supporting evidence from the Noun-Adjective transitions, this data does not seem sufficient on its own to support the Contrasting Structures Conjecture.

However, once again, the results of the native speaker group are a puzzle. They show the same sort of serial position effects as we found in Experiment 1, with the longest latencies lying in probe position 2, and there are no indications that the syntactic structure of the strings has any effect on the latency scores. The anomaly which we attempted to eliminate by revising the stimulus sentences has thus come to the surface again in this experiment.

Faced with consistently abnormal results of this sort, the obvious reaction would be to look for further explanations in terms of the stimuli used, or to search for effects which might be reasonably ascribed to the minor methodological differences which distinguish these experiments from other published work using the same experimental paradigm.

One possible explanation, for example, might be that the sentences used here are so short that they fail to produce the expected effects in the native speakers because they do not impose any strain on the short term storage capacity of the native speakers. With a small number of words in each stimulus sentence, a figure well within their normal storage capacity, it might be possible for native speakers to short-circuit any syntactic processing components, and to produce correct responses to a probe simply by using information already available in some buffer store that is normally used for rehearsal. This explanation seems unlikely, however, in that Suci, Ammon and Gamlin's original stimuli used only six words, and Kempen's sentences also contain only six words, one of which is an unprobed adverb. The materials used by these authors are not markedly different from the materials used in the main experiment reported in this chapter, which comprised sentences consisting of five words, yet both Suci, Ammon and Gamlin and Kempen report highly significant probe position effects that are clearly related to the syntactic structure of the input sentences. It is difficult to imagine that a reduction of one word would be responsible for such a fundamental change in the performance of the native speaker group - and in any case, explanations along these lines fail to account for the parallel performance



of the native speakers in Experiment 1, where the expected results also failed to materialize, despite the fact that sentences containing six words were used there.

A second possible explanation might be sought in the very restricted range of syntactic structures tested in these experiments. Here again, however, the sentences used were broadly comparable with those used in other experiments using the probe latency technique, and there is no obvious reason why these particular sentence types should give rise to unexpected anomalous results. The sentences used in both Experiment 1 and Experiment 2 were closely modelled on those used by Suci et al, in that they do not contain rare or exceptionally complex constructions. They represent a fairly common syntactic type in Spanish, and one which has a very similar distribution in Spanish as Suci's original sentences have in English i.e. we are not falling into the mistake made by Macnamara of comparing Passives in English and French, forms which are formally equivalent, but functionally and stylistically very different from each other (cf 5-2 above). The only serious changes made - the reversal of Noun and Adjective word order in Noun Phrases - were : dictated by the choice of Spanish as the language to be investigated, and there is no obvious reason to suspect that this minor alteration could account for such a major change in performance.

Other possible explanations which focus primarily on the various properties of the stimulus strings will readily spring to mind. It seems unlikely, however, that any of these could provide a wholly convincing account of the native speakers failure, given that these

same stimuli do produce the appropriate structure - related response latency patterns in the non-native speakers. This finding suggests that there is nothing inherently problematical with the stimulus material, and that it might be more appropriate to look in other directions for a plausible explanation of why the native speaker group fails to perform in the expected fashion.

## 6-7-2 Two Further Possibilities.

In this section two possible explanations for the native speakers' performance are considered. Both explanations are speculative, in that no evidence is offered to support them. Both explanations are radical ones, however, with far-reaching implications not only for the study of performance in a second language, but also for psycholinguistics in general, since they question some of the fundamental assumptions that underlie much of the research work that is currently being carried out.

A. One possible explanation of the native speaker results is that the population studied here is an abnormal one, in that all of them were living in exile, and used their native language in only a restricted set of environments and situations, chiefly domestic. The effects of living abroad for a prolonged period of time in a foreign language environment on performance in one's native language have never been investigated, though many expatriates readily admit that they are less fluent in their native language than they once were. Vocabulary and accent obviously suffer as a result of prolonged residence abroad. The effects of linguistic exile on the less obvious aspects of language behaviour, such as sentence processing are not known. There are, however, some indications that removing children from an immersion situation has some effects on their syntactic ability in a second language (cf Cohen (1975)) and it is not implausible to suggest that something similar might also happen in the case of adults who leave their native language environment. Constant and prolonged exposure to a second language might

cause some disturbances to the normal first language processing mechanisms, and one would expect such disturbance to show up in a structure sensitive task like the Probe Latency Test. Clarke (1976) has argued that immersion in a second language environment can cause a kind of "culture shock" akin to schizophrenia, and it is well-known that schizophrenic patients seem to have a general lack of sensitivity to syntactic structure in sentences (cf Maher (1972)), so that the performance of the native speakers might be ascribable to disturbances in their normal ability to process language. The speculative nature of this suggestion will be apparent, but if it could be shown that syntactic sensitivity was impaired by residence abroad - for whatever reasons - then this would have some serious implications for psycholinguistic research. Although the vast bulk of current psycholinguistics is carried on in English (cf Section 7-2-1 for a further discussion of this point), some work is being conducted using speakers of other languages, notably French, (cf. for instance Grosjean (1977)) and where experimental work is carried out on second language speakers, it is common to run a control group of native speakers as well. Most of this work is carried out in the USA, and for obvious reasons depends entirely on expatriate subjects. If subjects of this kind are untypical in the way they react to stimuli in their native language, the value of this work would be much reduced, and the reliability of its results seriously undermined. This is clearly an idea which needs to be investigated further with some urgency.

B. The second possible explanation of the native speakers failure to produce the expected results runs along rather different lines. One of the chief assumptions made throughout the work reported in the preceding chapters is that it is possible to generalize from the results of experiments based on the performance of native speakers of English on experimental material in the English language to the putative performance of native speakers of other languages on experimental material in their mother tongue. Thus, we have assumed that any structural effect found for English should also hold for other languages, and in particular for the two languages used here: French and Spanish. Generally, this confidence has not been misplaced: the native speakers of French tested in Chapter 2 showed a typical structure-related performance on recall of Statistical Approximations; the native speakers of Spanish studied in Chapter 3 showed larger click displacements for structured material than for unstructured material when these were presented in Spanish; and in Chapter 5, native speakers of Spanish showed clear lower thresholds for structured word-pairs in Spanish than they did for the unstructured ones. However, all three of these tasks are fairly crude, in that they rely on a broad contrast between structured and unstructured material, but do not examine closely the way native speakers perform on different types of syntactic structure. The probe latency task - which is potentially a more sensitive tool than the other three used - does do this. This line of argument suggests that one possible difference between native speakers of English and those of some other languages, including Spanish, is that while all are sensitive to the gross difference between totally structured and totally unstructured strings, sensitivity to the fine detail of syntactic structure may not be a universal characteristic of native speaker performance.

Two characteristics of Spanish make this a plausible argument, for though Spanish is normally considered to be a language that is not grossly different from English as far as its Syntax is concerned, it does show two features which are radically different from anything to be found in English, and which may have far-reaching consequences for the way native speakers of Spanish process the sentences of the language.

Canonical Sentence Structure in Spanish suggests that Spanish should be considered as a Subject-Verb-Object language (SVO) - i.e. the same basic type of language as English, In practice, however, word order in Spanish is much more flexible than in English, and it is not uncommon to find subject and verb inverted as in the examples below:

- 1a) mi padre compró un coche nuevo
- 1b) compró mi padre un coche nuevo
- ie my father bought a new car
  
- 2a) una mujer entró por la puerta
- 2b) entró una mujer por la puerta
- ie a woman came in through the door
  
- 3a) la mujer que Juan vio era mi hermana
- 3c) la mujer que vio Juan era mi hermana
- 3c) era mi hermonalla mujer que vio Juan
- ie the woman that John saw was my sister

Most reference grammars of Spanish devote little attention to this question. Harmer and Norton (1935) for example, merely state "The word order in Positive sentences is also very flexible in Spanish. The subject in most sentences may be placed after the verb without any change of meaning" (p59) and later "the subject of a principal clause or a subordinate clause is often placed after its verb..."(p506)

In fact there are some clear stylistic and semantic constraints on this inversion which restrict its occurrence, but in general the resulting Verb Subject Object word order is common both in written and spoken Spanish, and appears as an unmarked form in a wide variety of different registers. A full discussion of these constraints on this inversion which restrict its occurrence, but in general the resulting Verb Subject Object word order is common both in written and spoken Spanish, and appears as an unmarked form in a wide variety of different registers. A full discussion of these constraints is to be found in Green (1976). Green presents a very strong argument in favour of regarding the VSO order "with SVO as a topicalized alternative" as the basic underlying word order in Spanish, since the language behaves in an number of important respects as an example of the VSO languages described by Greenberg (1966).

A second major difference between English and Spanish is to be found in the morphology of verbs. Spanish has a very rich set of verbal suffixes, which vary according to the tense and aspect of their verb, and, more importantly, according to the person and number of the verb's subject. This means that there is a much closer relationship between verbs and their subjects in Spanish than is the case for English - indeed, it is very difficult for verbs in Spanish not to provide information about their subjects in the normal course of events. Consequently, where the subject of a verb consists of a Pronoun rather than a full Noun Phrase, the information it conveys is largely redundant, since it can be readily recovered from the Verbal Suffix. Normal practice in Spanish is for such pronouns to be omitted except where distinctions of emphasis are to be made.

Thus, in example 4 below, 4a is the normal unmarked form, while 4b is strongly marked, and would normally be used only in a sentence where some clear contrast was intended, such as 4c.

4a) Compró un coche nuevo  
he-bought a new car

4b) Él compró un coche nuevo  
HE bought a new car

4c) Ella no tenía dinero, pero él compró un coche nuevo  
SHE had no money, but HE bought a new car

Sentences such as 4a appear to consist of a lone Verb Phrase, with no overt subject, but a better analysis might be to look upon them as Verb (Subject) Object sentences, the subject finding its realization in the inflection on the Verb.

We have seen, then, that there is a strong tendency for Spanish to use a Verb Subject Object word order in sentences where subjects are overt, and that there is a strong in-built bias towards VSO structure in the morphology of the language. The main effect of both these characteristics is to enhance the importance of the verb in Spanish relative to other major constituents, and to severely reduce the importance of the Major Constituent Boundary. Truncated sentences such as 4a contain no main constituent boundary. Equally, sentences where the Verb and its subject are inverted do not have a clear distinction between the subject Noun Phrase and the Verb Phrase, as the Verb Phrase constituent has been broken up by the intruding Subject Noun Phrase. Indeed, Green's (1976) analysis of Spanish as a VSO language with SVO as an alternative for sentences with topicalization suggests that even in overt SVO sentences, the Verb Phrase

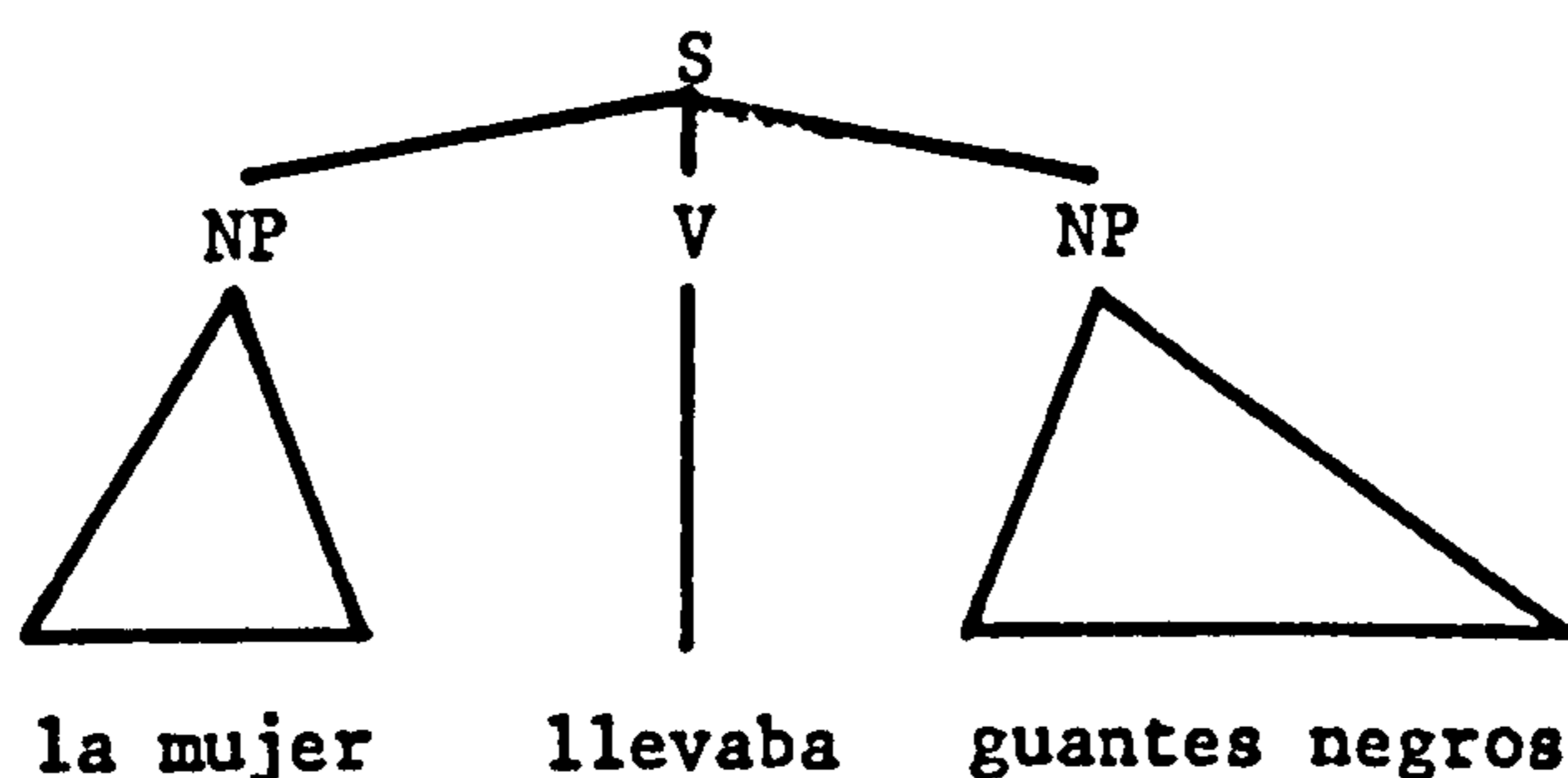


constituent arises as a result only of movement transformations, and may not play any major role in the syntax of the language.

The fact that word order in Spanish is much more flexible than in English, and the fact that this flexibility makes the constituent structure of Spanish sentences much less rigid and hierarchical than English, must mean that for native Spanish speakers the order of constituents in Spanish sentences is a far less reliable clue to the meaning of sentences than is the case for English sentences. Sentence Processing Strategies such as those discussed by Kimball (1973) will just not work for Spanish, since they rely too heavily on strict word order constraints. This makes it plausible to suggest that these characteristics of their language might have lead native speakers of Spanish to develop strategies of processing sentences which are relatively less dependent on the order in which constituents appear than the equivalent strategies used by native English speakers - and in particular that these strategies would not make use of the Verb Phrase notion as a central construct. If this were true, then there would be no reason to expect the major constituent boundary of sentences to have any psychological significance for native speakers of Spanish and the effective "psychological structure" of a sentence such as

la mujer llevaba guantes negros  
the woman wore black gloves

would be



rather than the standard analysis with the Verb Phrase constituent. An analysis along the lines shown above, would, of course, provide no reason for expecting the Probe Latency Task - or any other structure sensitive task - to produce markedly different results in the Major Constituent Boundary. There is no major constituent boundary which could do so.

The argument is very compelling. It is in fact doubly attractive, since if it could be shown that the sentence processing strategies of native Spanish speakers are systematically different from those developed by native English speakers, then the results of the learner groups studied in these experiments become less of an embarrassment. It will be recalled that these groups produced latency curves that were closely related to the syntactic structure of the stimulus sentences, or rather they were closely related to the syntactic structure that would be ascribed to a comparable English sentence. It could be argued, then, that the native English speakers produce the results they do because they are using processing strategies that are more appropriate for handling English sentences than they are for Spanish, and it is the use of these typically English processing strategies that gives rise to the pronounced Major Constituent Boundary effect. This would suggest that non-native speakers do not typically develop strategies for processing sentences which are akin to those used by native speakers, and where possible they use strategies which work for their native language.

These ideas are important, and clearly deserve further study.

Unfortunately, such work lies beyond the scope of this thesis. The sentence processing strategy notion has not been developed into

anything like a coherent theory, and there has been hardly any work undertaken on languages other than English which might allow us to explore these ideas further. The work reported in this chapter does not provide any clear indications as to how the sentence processing strategies of native Spanish speakers might differ from those of native English speakers, and a great deal of ground work would be necessary before any experimental work along these lines could be undertaken. Nevertheless, the indications are that further work along these lines might be very fruitful.

CHAPTER 7. CONCLUSIONS

7-1 Summary

7-2 Positive Findings

7-3 The Centrality of Syntax? A Reinterpretation.

7-4 Universal Psycholinguistics

7-5 Suggestions for further work.

7-1. SUMMARY.

In the first chapter of this thesis, we put forward what was apparently a simple, straightforward claim about some differences between native speakers of a language and learners of a second language. This claim, the Word-List Conjecture, suggested that learners handle foreign language material in what is essentially a word-by-word fashion, instead of processing incoming material in large, syntactically or semantically motivated chunks as normal native speakers appear to do. The subsequent chapters were concerned with testing predictions derived from this conjecture in a variety of experimental settings.

Four experimental paradigm were used: immediate recall of statistical approximations, click location, recognition thresholds for phrases and a probe latency technique. In each of these experiments, the performance of a group of fairly advanced learners was compared with that of a group of native speakers.

The result of these experiments confirmed that major differences between native speakers and learners do exist. However, not all the differences found supported the initial conjecture, and it was necessary at several points in the argument to modify the conjecture in the light of these experimental findings. In other words, this investigation has proceeded along the lines discussed by Popper (1968), in that we have deliberately attempted to find evidence that would refute our initial conjecture, derived new hypotheses from these refutations, and then examined these new hypotheses in more testing experimental situations.

The main findings are as follows. In Chapter 2 it was shown that non-native speakers produced very small, but none-the-less clear, structure related effects when they were asked to recall statistical approximations in their foreign language. Recall of higher order approximations was not reliably better than recall of low order approximations for the learners, but more sensitive measures of performance that looked at ordering effects in recall, and the size of sequences recalled as wholes rather than piecemeal, showed that there was a tendency for behavior to change in the same general direction as was found with the native speakers, but to a rather lesser extent. These findings are not wholly compatible with the predictions made by the Word List Conjecture.

A second set of experiments was designed to test whether the differences found in Chapter 2 could be ascribed to difficulty experienced by learners in organizing output of a complex kind, rather than to problems in initial processing. This work, using a click placement technique, was reported in Chapter 3. It produced results which it was not possible to interpret sensibly using the models generally considered as explanatory in this paradigm, and this particular line of approach was therefore abandoned. In the light of subsequent work, however, this failure can be seen to be of some importance, and its significance will be discussed below.

The refutation of the Word-List Conjecture discussed in Chapter 2 lead us to revise the Conjecture, and arguments were presented which led us to postulate that a distinction ought to be drawn between syntactic structures of a high level kind that consist of full lexical formatives, and a second lower level type of structure usually signalled by a grammatical formative serving as a marker of the constituent type. It was argued that this second type of structure might be expected not to

effect performance adversely, and that the Word-List Conjecture might still hold, as long as it was restricted to syntactic relationships of the first type. This argument was elaborated in Chapter 4.

Some evidence that supported the Word List Conjecture in this revised form was presented in Chapter 5. Here native speakers of both English and Spanish were shown to produce clear structuring effects in their native language, while failing to do so in their foreign language. Recognition thresholds for structured phrases in the native language were systematically lower than recognition thresholds for pairs of words with no structural relationship. In the foreign language, however, this difference was not found, and there was no evidence to suggest that structured material produced reliably lower recognition thresholds than unstructured material. This interaction between language and type of stimulus string was in line with predictions derived from the Revised Word-List Conjecture.

Inherent limitations on the recognition threshold type of experiment lead us to attempt some more searching tests of the Conjecture using a Probe Latency technique. The results of these experiments, described in Chapter 6 were a spectacular failure. Not only did they produce results that were totally incompatible with predictions derived from the Word-List Conjecture; they also forcefully suggested that some of the basic assumptions underlying this work are in fact untenable, and that a radical re-appraisal of the work reported so far is necessary.

This reappraisal is reported in the sections that follow. Section 7-2 begins by taking the findings at a fairly superficial level, treating them on their own terms, and attempting to assess the value of this work for what it tells us about the psychology of second language learning. In some ways, however, these simple findings are the least interesting thing to have emerged from this work, and of much more potential interest

are the doubts and problems about wider issues that these experiments have brought to light. One relatively minor methodological problem concerns the measurement of proficiency in second language speakers, and this is discussed in Appendix 7-C. Two other more serious questions, which have far-reaching implications for psycholinguistic research in general, are discussed in sections 7-3 and 7-4. Section 7-3 considers the question of interpreting the experimental methods used in this work, and suggests that they may not be measuring syntactic behaviour after all. This leads to a major re-evaluation of the results, and points to serious shortcomings in the standard interpretation of these experimental paradigms. Section 7-4 discusses further the question of a universal psycholinguistics. The final section, 7-5 considers some suggestions for further work, and in particular the advisability of using experimental methods to investigate questions in second language learning.



## 7-2. POSITIVE FINDINGS.

At the very simplest level of analysis, a number of positive conclusions can be drawn from the experiments reported in the preceding chapters.

The main finding is that the experiments failed to provide a totally convincing refutation of the Word-List Conjecture, but also to provide unequivocal evidence which could be interpreted as supporting it, and in this respect at least, the experiments are unsatisfactory. It would be a mistake, however, to give the impression that the work has been a total failure, for this is clearly not the case. All the experiments reported here showed highly significant differences between the native speakers and the learners. These differences were not always the ones that were predicted, indeed, some of the findings were totally unexpected, but none of the experiments reported here were inconclusive. The difficulty with them lies not so much with the findings as such, but with elaborating an adequate interpretation of what the findings might mean when they are taken as a whole.

Briefly, the new findings to emerge from this work are as follows:

a: in Chapter 2 we reported that non-native speakers of a language were less able to take advantage of syntactic redundancies than native speakers were. Though there was some evidence of their attempting to structure their output differently with higher order strings, the overall effect of order of approximation on the performance of learners was negligible. Furthermore, whereas the native speakers, though

scoring proportionately worse on the long strings, increased their scores in absolute terms, the learners failed to do so, and their raw scores on the 20-word strings were not significantly different from their raw scores on the 10-word strings. This finding suggested that non-native speakers may be suffering from a very severe limitation on memory capacity which native speakers are immune from. The reasons for this difference are not at all obvious, though a possible explanation is discussed in section 7-2-3. The practical implication of this finding is that if one wanted to investigate further the nature of differences between native speakers and learners in their language behaviour, a good place to look would be tasks which involved long sequences of input and relatively heavy memory demands. The severe limitations hinted at here suggest that ordinary conversations (i.e. not "foreigner talk") might, with ingenuity, be made to serve this purpose.

b: Chapter 5, the only one which provided clear evidence in support of the Word-List Conjecture, showed that structuring in short phrases consisting entirely of full lexical words, fails to have an effect on the performance of learners of either English or Spanish in a tachistoscopic recognition task. Since the task involved here is relatively straightforward, the indication is that further investigations on rapid reading tasks might be a fruitful source of further differences between native speakers and learners.

This task is perhaps the most naturalistic of all the experiments used in this thesis, and though still experimental in character, it is the one least far removed from real language behaviour in everyday life.. This clearly highlights the importance of the finding, as there is a higher degree of face validity in the experiment than in some of the others.

c: Chapter 6 also showed marked differences between native speakers and learners of Spanish in their response to a probe latency task. The non-native speakers were consistently slower in their responses than the native speakers were. In addition, important qualitative differences in the pattern of responding emerged. Some possible explanations for these differences and what they imply have already been discussed.

d: Additionally, in Chapter 3, we showed that subjects operating in their second language were considerably less accurate at locating the objective position of a superimposed click than native speakers are, even when the material consists only of isolated words and digits. The full implications of this finding were unclear, however.

In order to assess these findings fairly, it is of course, important to place them within an appropriate context. Experimental work only makes sense within the framework of the context that generated it, and the work reported here is clearly no exception to this general rule. Experimental studies of foreign language learners are relatively rare, and what work is carried out is perhaps best seen as belonging to what Kuhn calls the pre-paradigm state that precedes the development of "normal science". Inevitably this sort of period involves research which sometimes takes one in directions that were unexpected. Given the very limited state of our knowledge about second language learners and what they do with their additional languages, unexpected findings such as those reported here which were not all predictable in any obvious way from what was known before, make an important contribution to our understanding of the underlying problems, if only because they make these underlying problems rather more easy to identify.

With hindsight, it is perhaps not entirely surprising that the empirical data obtained in these experiments appears to be inconsistent, if not downright contradictory. Though each of the experimental chapters contained a section in which previous work with second language learners was discussed, it should be readily apparent that as far as the experimental techniques used here are concerned, at least, hardly any systematic study of foreign language learners has been undertaken. Considering the vast numbers of people who learn foreign languages, surprisingly little is known about how non-native speakers behave when faced with foreign language material, and the state of our knowledge in this field compares unfavourably with what we know about smaller and far more esoteric groups who have been intensively studied by clinical psychologists, for example. Much of the work that has been undertaken in the study of second language learners is unsatisfactory on a number of counts. A lot of this recent work is merely a reworking of old ideas (cf Cook 1980) which adds little by way of new data or new interpretations. A great part of the work in the field is, unfortunately, experimentally crude - Macnamara's work for example, which was extensively discussed in Chapter 5 is linguistically insensitive and omits even the most obvious controls on the stimuli, as well as using dependent variables which contain a high level of inbuilt and unavoidable error. Other work, highly sophisticated at first glance, actually turns out on closer examination to be based on assumptions that can readily be shown to be false. Triesman's work, discussed in Chapter 2 illustrates this problem. Her assumption that entropy values for French passages would be identical to the entropy values of English passages of the same order of approximation seems totally unjustifiable when one takes into

consideration the major syntactic and morphological constraints that distinguish English from French. Once this assumption is queried, it becomes impossible to accept Triesman's regression equations at face value, and this in turn seriously undermines her discussion of the differences between native speakers and learners, and invalidates her conclusions.

Given such an inadequate level of background research, it would be unreasonable to expect clear and unequivocal conclusions to emerge from what must inevitably be seen as an exploratory study. The more realistic aim - that the experimental data should point to inadequacies in the initial formulation of the questions being asked - is certainly one that has been achieved, however. Probing of the kind carried out here should make it possible for more carefully formulated questions to be asked by future researchers, and at the very least, this work has served to pinpoint a number of interesting questions which could easily be examined and tested, using a wide range of experimental tools and research methods.

This assessment still leaves us with the need to provide some principled account for why the data turn out the way they do, however. Clearly, our original attempt at an explanation, that learners differ from native speakers in that they are not sensitive to syntactic structuring, is not going to be able to account for the results, since, as we have seen, some of the experiments described produce results which are not readily interpretable within a framework of this sort. On the other hand, the fact that the data presented here, while apparently unsystematic, actually corresponds closely to the sort of

behaviour produced by other non-standard populations, suggests that there ought to be a simple explanation of the data, and a single principle that underlies them all and could make sense of the disparities we have reported. A detailed discussion of the similarities between these data and comparable data from children and schizophrenics will be found in Appendix 7C. These similarities are not in themselves, of course, a sufficient explanation for the behaviour of the learners studied here, nor does the existence of these similarities imply that they have identical or even similar causes. A full examination of this latter question is clearly one that lies outside the scope of this thesis. Nevertheless, the similarities are sufficiently interesting for it to be worthwhile making an attempt to impose some kind of order on the data, rather than just dismissing it as a series of unrelated and contradictory phenomena, and an attempt to do this will be made in the next section.

### 7-3. THE CENTRALITY OF SYNTAX. A RE-INTERPRETATION.

One of the fundamental assumptions on which this work has been based is the assumption that the experimental methods used were reliable tests of behaviour induced by syntactic variables. When the work was first begun, it seemed reasonable to believe that the various structural effects reported in the experimental literature were a unified set of phenomena which all arose as a consequence of the way syntactically structured material was handled by native speakers of a language. This belief was encouraged by the work of Fodor, Bever and Garrett (1974) who report the results of a wide range of structurally based experiments as though they formed a single coherent corpus of data. This work implied that the many different sources of evidence supporting the claim that syntactic structure is "psychologically real", far from being a weakness, is actually a strong argument in its favour, since the claim does not rely exclusively on evidence obtained in a single experimental paradigm. In the absence of a properly worked-out and comprehensive theory of language processing, it seemed economical to believe that one of the principal steps involved in handling sentences involved the computation of their syntactic structure, and that once computed, this internal representation of syntactic structure might be expected to manifest itself in suitably designed experimental situations. The strategy used here was to show that the overt manifestations of syntactic behaviour found in experiments using non-native speakers as subjects differed systematically from those found with native speaker subjects, and to infer from this that the second language speaker's

covert representation of the syntax of sentences differed in a predictable way from what would be expected of the native speaker. It would follow that the second language speaker's syntax processing abilities were defective in some way.

The principal difficulty with this argument is that it assumes that the syntactic effects studied are pure - that is, that one can devise experimental settings which effectively isolate the syntactic processing stages of speech perception from all the other processes involved in handling language. There might be some justification for this assumption in the case of native speakers of a language - though even here the current trend in processing models is for a method of parallel processing, which implies that any experiments that did successfully isolate syntactic functioning would necessarily be very artificial and remote from real language behaviour. In the case of learners, however, it is much less clear that there is any justification for an assumption of this sort. In fact, the work reported in the previous chapters has repeatedly thrown up hints that the assumption is an invalid one.

Encoding syntactic structures and performing even a simple operation on these forms is a task of some complexity, and involves a number of important subskills. Normally, however, it seems reasonable to assume that these subskills can be ignored for experimental purposes. They are components of any verbally-based task, and can therefore be assumed to operate in a more or less constant manner, effecting performance in all sorts of verbal material used in experiments in much the same fashion, and to a comparable extent. Thus, though these subskills



are a source of some noise in experimental data, they are a source that can safely be ignored in a large number of experimental settings. Consider as an example of an important subskill of this sort, the skill of word recognition. All verbal tasks involve the recognition of words, and yet word recognition is largely ignored in experimental work. Where it is essential, for most practical purposes the variations in the data due to differences in the ease with which different words are recognised can be reduced by introducing some simple controls for obvious factors such as frequency, length, pronounceability, and so forth. The remaining variation would be expected to be small relative to the size of any effects due to the experimental variables under investigation. In fact, recognition problems account for such a small proportion of the variance of experimental data in studies of syntactic processing, that even these elementary controls are often omitted, and for all practical purposes, word recognition is treated as an instantaneous phenomenon. The whole literature on click placement, for example, contains no instances of controls of this kind being used.

However, if we turn now to the experiments reported here, it soon becomes apparent that there is a marked language effect in all the experiments. Performance with material in the second language of the subjects is consistently poorer than performance on tasks where the material is presented in their first language, irrespective of whether this material is syntactically structured or not. This suggests that those fundamental subskills that function automatically and painlessly for the native speaker may not always do so for the second language speaker. In fact, in all the experiments reported here, the language effects and the interactions between language and group are generally much more

significant than the effects of the syntactic variables which were the principle concern of this thesis. This makes it seem likely that difficulties with syntax, real though they obviously are, may actually be far less important for the learner than some basic deficiencies in his ability to handle second language material that we have not attempted to study.

By way of illustrating this point, consider further the process of word recognition, the most obvious and most accessible of these fundamental subskills. Let us concede that it might be permissible, and even desirable, to ignore word recognition in native speakers on the grounds that the process is effectively instantaneous, and that variations in word recognition contribute only a small fraction of the total variance of the scores in a syntax-orientated study. The question which now arises is whether word recognition processes can be safely ignored in this way where second language speakers are concerned. The answer to this question may be yes, probably, if it could be shown that word recognition in a second language took a constantly greater length of time than word recognition in a first language, or was consistently greater on any measure of difficulty than word recognition in a first language; or indeed, if any other easily definable function could be found which provided a mapping between performance in a first language and a second language, then the answer would be affirmative. There are, however, strong grounds for believing that no such function exists.

In the section that follows, we shall discuss in detail some of the ways in which the formal properties of words differ across languages. For the moment, however, a few brief comments should serve to make clear the nature of the problem. Consider, for example, the fact that

languages differ in the distribution of word length among their lexical items, in whether words are inflected or not, and if so where these inflections occur, and in whether words are generally highly distinctive, or very similar to other words in the language. It seems likely that formal differences of this type would affect the non-native speakers ability to recognize words in much the same way that contrasts in syntactic patterns tend to produce difficulties for non-native speakers. Thus, a speaker of a basically mono-syllabic language might be expected to handle, say, English words relatively easily, as long as he was limited to monosyllabic words, but he might be expected to experience some considerable difficulty with polysyllabic words, even when these are not an obvious source of difficulty for the native speaker. Similarly, a native speaker of a language that is predominantly polysyllabic, and contains a high proportion of three or four syllable words among its most frequent items, might find the very large number of highly frequent English monosyllables disturbing and difficult to discriminate. Formal differences of this type would mean that words that were easily recognised by native speakers could be numbered among the most difficult for a given group of non-native speakers. Quite the opposite problem is encountered with highly infrequent words in English which are often of Greek or Latin origin, and have relatively high frequency cognates in other languages. In English, these words tend to be restricted to a rather narrow range of learned registers, but their cognates in the Romance languages and Greek are often in common use. If we assume that cognates are handled with ease by non-native speakers, then it seems probable that many infrequent and learned forms in English would be much easier for the second language speaker to handle than their highly frequent everyday English equivalents. In short, then, it seems unlikely that there exists any one-to-one mapping between the rank order of difficulty of English words for a native speaker and the rank order

of difficulty of the same words for a non-native speaker. This makes it unwise to ignore word-recognition as a factor in experimental work involving non-native speakers, but at the same time renders useless the normal controls that one would use in order to minimise the effects of variables of this kind.

The frequency problem alluded to above is actually even more important than it appears at first sight, for even if the other practical problems were solved, and materials which were not formally easier for one group could be constructed, the frequency problem would still remain. Given that the second language learners studied here had not lived abroad for any length of time, and that exposure to the target language was essentially limited to classroom experience, it seems unlikely that even the most frequent items of vocabulary in the target language were actually frequent items when viewed in terms of their total stock of words in both the target language and their native language. Much more likely is that even the most frequent items in the foreign language would have a subjective frequency of occurrence corresponding much more closely to that of a relatively infrequent word in the native language.

Now, we know that infrequent words have quite different psychological properties from frequent words, and that frequency interacts with syntactic variables in some experiments - infrequent words have a tendency to produce syntagmatic word associations in adults, for example (Stoltz and Tiffany (1972)). This means that even when a frequency control is used, as we did in these experiments, the comparison of native and non-native speakers is still by no means as straightforward as it looks at first sight. In effect, one group is being tested with words of a low frequency, while the other is being tested with high frequency words, and under these

circumstances, obvious assumptions about the advisability of ignoring word-recognition as a factor in syntactic experiments may clearly no longer hold.

Basically, then, we have been assuming throughout this work that the syntactic tasks studied here are essentially the same for both the learners and the native speakers, but this assumption now appears to be naive and simplistic.

The obvious question that arises at this stage concerns whether or not it is possible to make sense of the experiments reported here in terms other than syntactic ones, and in particular, whether the arguments advanced above about word recognition are of any help in this respect.

In fact, two of the sets of findings can be very easily explained in terms of the word processing capabilities of the non-native speakers subjects. These are the results of the click-placement experiments, and the tachistoscopic recognition threshold experiments.

Firstly, we have already shown in Chapter 3 how a set of apparently nonsensical and anomalous results could be readily interpreted if it was assumed that non-native speakers took slightly longer to recognize words than native speakers did. At the time, we argued that this explanation probably destroys the credibility of the click-placement task; however, it may be the case that word recognition is basically what accounts for the incorrect placement of clicks, rather than syntactic structure, and what appear to be genuine structural effects in the behaviour of native speakers may be produced only because the presence of syntactic structure facilitates word recognition. This clearly is an instance of a problem to which we shall return in section 7-5 viz

the difficulty of finding reliable interpretations of phenomena turned up in experimental settings.

Secondly, the experiments reported in Chapter 5 are also amenable to an interpretation in terms of word recognition. The question asked in that chapter was whether context facilitated the recognition of words in a foreign language, and the answer to this question was a clear negative. It is presumably not without significance that this question, the one that refers most clearly to word recognition, was the only one that provided a wholly unequivocal answer of this sort.

The two remaining experiments are less easy to explain satisfactorily in terms of word recognition, simply because the tasks that they examine are conceptually more complex than those of the other two chapters, and it is correspondingly more difficult to establish at what point the learners' behaviour diverges from that of the native speakers.

The experiments on statistical approximations were simple memory experiments, involving immediate recall, and since the words were presented at a slow rate, and in good acoustic conditions, slower word recognition ought not to have affected performance. Nevertheless, it is just possible that here, too, failure to recognize words might actually be responsible for the lack of expected syntactic effects in the learner population. Two lines of argument lead to this surmise. Firstly, if learners failed to recognize any of the words read to them, then they would automatically fail to make any of the syntactic links between that word and adjacent words. This possibility was controlled for by checking that students did know the words presented, but it is possible that subjects' self-reports in such situations are unreliable, or that they

did indeed recognize words in the post test, but failed to recognize them when they first heard them.

Evidence from Word Association tests and from observational studies of malapropisms in learners suggest that learners frequently misunderstand words, by confusing them with other similar sounding words in the foreign language. They appear to be almost completely unaware of this problem. (cf Meara (1978)) A second line of argument goes something like this. Suppose the learners are theoretically capable of recognizing the words used in the experiment. Imagine that words are presented at a rate of one every 0.n seconds, and that the average time taken to recognize a word varies according to certain unspecified properties (e.g. length, complexity, and so forth) but that this time is fairly close to 0.n seconds. Now, if a hard word is presented, there is a possibility that it will not be fully processed before the next word is presented. Effectively, then, the gap between words is nullified, and this could have two further consequences: a) there is no time for rehearsal of the last item learned, and b) the processing of the last item is likely to interfere with the processing of the next word in the series, and so to make it less likely that that item is fully processed before the following one is presented. If a relatively long average word recognition time such as we have described here did cause mis-recognitions, then again one would expect the normal syntactic effects not to materialize. These considerations illustrate very clearly the difficulties one runs into when one assumes that experimental tasks are identical for learners and for native speakers. With complex tasks of this sort, it is quite probable that the predicted results might emerge for reasons quite different from the obvious ones.

The results of our probe-latency experiments are rather less easy to explain in terms of word recognition phenomena. This is partly because these experiments were problematical in any case because of the unexpected performance of the native speaker group, and partly because the nature of the task is such that it is more directed towards production than perception, in that it is the latency of the response that is the crucial variable. This latency is obviously made up of at least two sub-components, one of which is the recognition of the probe word, but given that this word is only one out of a possible five, it seems unlikely that this particular component will vary greatly. However, there is one interesting possibility concerning word production that needs to be taken into account. Basically, the longest latencies of the learner group are produced when the subject Noun is given as probe, and the complementary verb is the required response. Now it seems reasonable to assume that native speakers treat Verbs as single-word units, but it could be argued that teaching methods that stress roots and the affixes that are attached to them might lead non-native speakers to treat Spanish verbs as morphologically more complex than nouns, and consequently to take longer to produce them. Such a reading would suggest that the learners' behaviour on this experiment might not be syntactically induced at all, but rather a function of the morphology of Spanish words in contrast to their English counterparts.

All the explanations advanced in this section are clearly post hoc - though not wholly ad hoc - and are very largely speculative. Nevertheless it is not without importance that the experimental data we have presented here are open to interpretations of this sort, which are often wholly removed from the normal terms of reference used to discuss findings from



similar experiments using normal native speakers of English. The obvious inference from this is that our attempt to elucidate the role of syntactic structure in the performance of second language speakers was essentially premature, and before further work of this type is attempted again, it seems crucial that a much clearer picture of word recognition and word production in a second language should be established.

#### 7-4. UNIVERSAL PSYCHOLINGUISTICS.

The second assumption on which this work was based is also, in essence, a fairly simple one. It was believed that experimental work in psycholinguistics had established a number of effects that could be related to the syntactic structure of the sentences used in the experiments, and that these structural effects were a general property of the way native speakers of a language react to its sentences. It was assumed that these properties were not language specific, and that it would be possible to use results based on data culled from experiments using English to make inferences about how native speakers of other languages would behave when presented with comparable material in their own native language. However, the results of the experiments reported in Chapter 6 strongly suggest that this assumption may be false. In those experiments, the native Spanish speaking group failed to perform in the manner predicted by the models that are generally used to account for the Probe Latency effects found with native speakers of English. Some possible explanations of this finding were discussed in Section 6-7, and it was argued there that speakers of languages which have structural properties that are markedly different from those of English sentences may develop processing strategies for sentences in their own language that bear little resemblance to the processing strategies commonly found in experiments using native English speakers and materials in the English language. If this suggestion contains any truth, then it is clearly no longer possible for us to hold our initial assumption.

It is easy to see how and why one might fall into the trap of making an assumption of this sort. Much of modern linguistics is concerned with

defending the view that despite the many and varied superficial differences between them, all human languages are in a sense fundamentally the same. A large part of linguistic research is directed towards the task of establishing whether there are any universal properties which can be shown to hold for all natural languages, or at least for large subgroups of natural languages, and it is widely believed that such linguistic universals do exist, (cf. for example, Greenberg (1966), and Bach (1974), esp Ch. 11) Given a climate of opinion that is strongly concerned with formal linguistic universals, it is only a short step to a corresponding metatheoretical belief in psycholinguistic universals. If all languages have similar syntactic properties, then it seems plausible to assume that these properties will be handled in a broadly similar fashion by native speakers of any language, and it seems reasonable to imagine that it might be possible to prove the existence of a small set of processing strategies universally applicable to all languages, and parallel with the linguistic universals found in syntax. There is no obvious motivation for us to assume that there might be radically different processing procedures associated with distinct languages. This belief finds some support from the only area of psycholinguistics which has paid any real attention to the behaviour of speakers of 'exotic' languages - studies of the development of language in young children. In this field, thanks mainly to the influence of Slobin, a large number of studies covering a wide range of exotic languages has been undertaken - a situation which, as we shall see, is in marked contrast with research carried out in other areas of Psycholinguistics. The results of this work tend to support the view that psycholinguistic universals of some sort do exist, since the normal course of language development in

different linguistic communities shows a remarkably high degree of similarity, given the diverse nature of the languages studied. (cf. Ferguson and Slobin (1973) and Slobin (1970).<sup>18</sup>

However, it could be argued that studies of language development in children are to a large extent orthogonal to other areas of psycholinguistics. Typically studies of language development concentrate on the order of acquisition of the forms and structures that make up a language - i.e. they have centred mainly on the task of furnishing an account of how the child's linguistic competence develops (though Peters (1980) represents a rather more interesting approach.) Given that we already know there is a large degree of formal similarity between languages, it is not really surprising to find that these forms emerge in a sequence which, while not identical in all languages, is sufficiently similar to allow us to make meaningful comparisons across a wide range of languages. However, competence data such as this have no bearing on questions of performance such as the ones we are interested in. It does not follow that formal resemblances between languages, or similarities in the order of emergence of these forms, compel us to believe that all aspects of human linguistic behaviour will turn out to be equally comparable. In fact, given that processing strategies of the type we are discussing are predominantly concerned with surface-structure features of language and given that surface structures are the level where languages differ most markedly, there is every reason to subscribe to the contrary belief that normal language behaviour in one language might be expected to be markedly different from normal behaviour in another language - at least in those activities where surface structures are crucially implicated.

Unfortunately, this idea has never been systematically put to the test. Modern psycholinguistics has been characterized by an almost total dependence on experiments conducted in English. Only a handful of other languages have been studied at all, and these are mainly languages closely related to English. Experimental work on languages that lie outside the Germanic branch of the Indo-European family is negligible.

The full extent of this dependence on data taken from English language experiments can be seen from a cursory glance at any of the standard text books, but a particularly good example, which clearly illustrates the attitudes commonly held, will be found in Fodor, Bever and Garrett's (1974) Chapter 5. This chapter deals with a topic which is central to the concerns of this thesis - "the Psychological Reality of Grammatical Structures" (p221). All the data cited comes from experiments in English with the solitary exception of some work carried out in Dutch by Levelt (1970). Dutch is a Germanic language, closely related to English in its morphology and its lexical structure, as well as in its syntax, and so it is not surprising to find that Levelt's word-sorting procedure gives rise to hierarchical pattern structures that closely parallel those found using the same technique with English material. It is interesting to note, however, that Fodor, Bever and Garrett nowhere spell out the full implications of using data derived from Dutch in an argument that is otherwise concerned with the performance of native English speakers, and it is clearly implied that the language of the stimulus string is not considered to be a factor of any importance. Indeed, Fodor, Bever and Garrett discuss the stimuli as if they had in fact been English sentences, even replacing the original Dutch words by their English translations in

their account of Levelt's work:

"For example, the scaling structure for (5-13) is given in figure (5-8). The untidiness of this structure can be seen

5-13 Carla takes the book and goes to school

by noting that the verb in the second conjunct is as closely related to the first noun as is the verb in the first conjunct . . . . . In (5-13) the single surface structure occurrence of the word Carla was judged by Levelt's Dutch subjects as almost equally related to the two surface verbs takes and goes." (p254)

One might be forgiven for taking this as an account of an experiment in English which used native Dutch speaking subjects.

In this particular instance, there may be some justification for this insensitive handling of non-English data. Levelt's sentences are indeed very like their English translation equivalents, matching morpheme by morpheme on a one-to-one basis. This very similarity, however, sharply emphasizes the extreme narrowness of the data base on which Fodor, Bever and Garrett's widely accepted conclusions are founded.

Two major drawbacks arise directly out of this overreliance on English, which appear to have serious implications for the development of Psycholinguistics.

The first of these drawbacks is that researchers in psycholinguistics have been led to concentrate their efforts on problems which in the context of English seem to be important, but which can be readily seen to be only minor problems when viewed in a wider context. One example of this trend can be found in the enormous amount of work that was carried out in the late sixties and early seventies on the Passive Construction in English. While the results of this work are by no means trivial, it

is surprising that a syntactic construction which is in some ways quite language-specific should have acquired the status of a major theoretical concern by being studied extensively in this way. English passives are particularly complex compared to equivalent forms in other languages, and even where there is a close formal resemblance between English passives and those of other languages - French for example - there is often a marked difference in the distribution of Passive forms in ordinary language use. Considerations such as those suggest that it is improbable that the findings of research based on English Passives would be widely generalizable to other languages, and yet contemporary accounts e.g. Greene (1972)) seem content to accept such data as relevant to universal claims about the way human language is handled. A further example of this kind of problem will be found in the work which studied the effects of deleting relative pronouns from sentences. This property of English sentences is one which has been widely studied in a variety of experimental settings (e.g. Fodor and Garrett (1967), Bever and Langendoen (1971) etc. principally because it is one of a number of optional "stylistic" transformations which have the effect of making the surface structure of sentences less transparent, and, not surprisingly, renders them hard to process. Again, however, deletion of relative pronouns is not a common syntactic phenomenon, and is rare even in the Indo-European languages that closely resemble English in other respects - cf. Peranteau, Levi and Phares (1972). Fodor, Bever and Garrett (1974) cite this data as evidence in support of the speculation that linguistic universals may be "explicable by reference to the computational procedures that speaker-hearers use to process sentences" (p361), but the very restricted nature of the evidence discussed is heavily at odds with the universal emphasis in the claim.

The second major problem that arises out of this almost exclusive concentration on English is that it fosters the growth of an orthodoxy which may in the long run turn out to be essentially language specific, even though it is commonly treated as if it had a universal validity. In this way, constructs which can readily be shown not to apply in a wide range of languages come to assume a central role in the theory, which is thus reduced in its scope and its generality.

One good example of this is the current concern with the psychological effects of major constituent boundaries - that is the effects of the Noun Phrase - Verb Phrase segmentation commonly used by Transformational Grammarians in their descriptions of languages. The major constituent boundaries are an obvious feature of English and of other SVO languages. There are, however, no major constituent boundaries in the sentences of languages such as Welsh or Irish, which have a basic VSO structure (cf. Awbery (1976)) - nor, as we have seen, can we be certain of finding major constituent boundaries in the sentences of a language such as Spanish which is SVO in theory, but often not so in practice. Though linguists have clearly recognised these phenomena, psycholinguists have often failed to do so. The full implications of differences such as this are by no means clear, but it is not implausible to imagine that the lack of a characteristic as fundamental as a major constituent boundary might lead VSO languages to produce a psycholinguistics which radically different from the psycholinguistics of SVO languages.



Or again, in a similar vein, consider the semi-formalized processing strategies that are now being advanced as partial models of sentence handling. Fodor, Bever and Garrett's (1974) strategy 6-5 states:

"Take a verb which immediately follows the initial noun of a sentence as the main verb, unless there is a surface structure mark of an embedding." (p356)

This strategy will normally succeed for English, but it will fail immediately when faced with a VSO word order for quite trivial reasons, and the strategy fails to cope even with the subjectless verb forms that occur in Spanish.

Similar considerations apply too at the lower levels of processing. Here most of the strategy types that have been proposed rely heavily on function words as a way of identifying constituents (a point already discussed in Chapter 4). Kimball's (1973) NEWNODES strategy, for example, uses function words explicitly in this way:

"NEWNODES:

The construction of a new node is signalled by the occurrence of a grammatical function word." (p32)

Untypically, Kimball appears to be aware that this strategy will be inappropriate in many languages, and in a footnote to NEWNODES he adds:

"The operation of NEWNODES in SOV languages needs further examination. In such languages, grammatical formatives typically follow these constituents to which they are attached .... for large constituents such as sentences with following complementizers, NEWNODES is simply inoperative." (p33)

Kimball does not however attempt to elaborate any alternative strategy which could identify constituents for SOV languages or for other languages where function words would fail to signal new constituents. Nor does he consider the metatheoretical implications of the need for alternative strategies to be devised.

Given the very large number of languages which do not conform to the SVO word order, it is surprising to find that more attention has not been given to the study of general approaches to sentence handling. The strategies approach is much more language specific than it appears to be at first sight, and it is unlikely that a more genuinely universal theory of sentence handling will emerge as long as research centres on English and ignores other languages.

The field of syntax is not, of course, the one one where this reliance on data taken from English is apparent. The problem is a pervasive one, and gives rise to similar difficulties in most areas of psycholinguistic research. As a further example, let us briefly consider current theories of Word Storage (cf. Fay and Cutler (1977)). There is a general consensus, based on work such as Bruner and O'Dowd (1958) and Brown and McNeill (1966) that in the mental lexicon the phonological entries for words are incomplete, and that certain features of the surface phonology of words are more readily accessible than others. Brown and McNeill argued that even when people were unable to produce a current phonological form for a dictionary entry, they could often produce partially correct reports of the word they were seeking. In particular, subjects could often say how many syllables the target word contained, what the stress pattern of the target word was, what sounds its initial segment contained, and whether it included any infrequent consonant clusters. Bruner and O'Dowd showed that misspellings in the beginnings of words had serious effects on recognition, while similar spelling mistakes at the end of words had a much less marked effect and a misspelling in the middle of a word often passed unnoticed. They concluded from this evidence that the first syllable of a word is of crucial importance to its identification, while endings play a less important role, and the middle segments of words are almost wholly redundant. Data from spoonerisms, malapropisms and other slips of the tongue tend to support these claims. (cf. Ellis (1979)).

It is very common for findings of this sort to be discussed in universal terms, as shedding light on "the structure of the mental lexicon", and yet it is immediately obvious that none of the features discussed above would be likely to find its way into a list of universal characteristics describing lexical structure. It is, in fact, very easy to find languages in which the constraints that govern the structure of words make these features either uninformative or of marginal importance. At the simplest level, there are a large number of languages - such as Vietnamese - whose morphemes are essentially monosyllabic. With languages of this sort, the number of syllables in an item is wholly predictable, and thus does not discriminate between morphemes. Stress pattern is clearly irrelevant in such cases. (One might predict, however, since many monosyllabic languages are also tonal, that speakers of such a language might be able to recover the tone of the target word in a tip-of-the-tongue situation, although tone does not usually figure in current theories of word storage.) Even in Indo-European languages, the commonly accepted list of features will not always be illuminating. French words, for example, are all stressed in their final syllable so that here again, stress pattern is predictable and fails to discriminate between words as it does in English. Even the psychological importance of the beginnings and endings of words is in doubt in some languages. In Irish and Welsh, and the Celtic languages in general, certain grammatical and phonological environments can induce changes in the initial consonants of words ("mutations") which in theory ought to make them extremely difficult to recognize, but which in practice does not seem to do so. The endings of words in Romance languages are to a large extent redundant, as they contain mainly grammatical information about gender and number, and not lexical information, and in some languages

like Guarani, which have a complex agglutinative morphology, grammatical affixes appear both before the main lexical stem and after it in the word, and not in one of the more psychologically salient places. A completely different type of problem arises when one considers the structure of words in Semitic Languages. Here there is a complex interaction between consonant frameworks and vowel infixes which provide variations on the theme signalled by the consonant structure. This sort of lexical structure is quite unlike anything that occurs widely in the Indo-European languages, and it seems plausible to suggest that Semitic speakers would thus develop word processing and storage patterns which differ markedly from those found with native English speakers.

There seems, then, to be a strong case for arguing that what are often considered to be universal properties of word storage in the mental lexicon are actually highly language specific, and that as such they may not throw very much light on the processes of word storage and lexical access in languages other than English. If pertinent universals do exist, it is unlikely that they could be expressed simply in terms of the surface phonological structure of English words. Such putative universals are much more likely to take the form of general principles, rather than the particular instantiation of these principles in any one language. Again, however, such general principles are likely to remain obscure as long as research concentrates heavily on English.

## 7-5. SUGGESTIONS FOR FURTHER WORK.

The two preceding sections have discussed two serious problems which have arisen in the course of this work, and some of the implications of these problems for psycholinguistics in general. At this point in a thesis, it would be normal to outline suggestions for further research work along the lines developed earlier, perhaps to draw attention to a number of minor problems which remain to be tidied up, and to point out any practical applications of the findings. In this case, however, such an approach is not altogether satisfactory. None of the findings has any obvious practical applications, and the main thing to have emerged from this work is a cluster of uncertainties in an area where previously there were few. Raising problems in this way is an important part of research activity, of course, and the identification of problem areas is an important preliminary to any further work, albeit a rather unsatisfying one.

For those who remain undaunted by the very serious problems that have been thrown up by this attempt to use experimental methods with second language learners, three directions for further research seem to present themselves.

a: For those who still believe that it is desirable to carry out experimental studies of language learners, there is clearly a very strong case for finding out much more about the basic processes of word handling in a second language. We have already seen how word handling abilities may provide a key to understanding the anomalous

results found in these experiments, and how almost all other types of experiment that make use of verbal material may be failing to make an obvious control by assuming that word recognition in a second language is in practice indistinguishable from word recognition in a first language. The difficulty with this line of approach is that current experimental paradigms for investigating the handling of words either concentrate on techniques which look at isolated words, and might, therefore, be criticised for being far removed from the world of real behaviour, or else they concentrate on the processing of large scripts or texts, and thus run into some of the problems of inference that we discussed above. It is not clear what would count as a satisfactory solution of this dilemma.

b: At another level, there is clearly a need for a series of experiments attempting to explore the ideas developed in Section 7-4, where it was argued that superficial differences between languages could give rise to a psycholinguistics that might be radically different from the psycholinguistics of English. This line of approach is basically concerned with questioning the standard orthodoxy in psycholinguistics, by disputing whether the basic findings of the discipline are genuinely universal properties of the way humans handle language.

c: The third direction pointed out by these experiments is also concerned with challenging the orthodoxy, but does so rather more on the level of interpretation of the findings, than at the level of the findings themselves. This avenue of research would be concerned principally with testing a wide range of alternative explanations for many of the standard findings in psycholinguistics. One of the main

things to emerge from the experiments carried out here is that the standard models used to explain a wide range of psycholinguistic phenomena really do not make very much sense when they are applied to the results of experiments with foreign language learners. This suggests that despite the wide agreement over these interpretations to be found among psycholinguists, experiments of this type are in reality only poorly understood, and probably widely misinterpreted. This must raise serious doubts about the wisdom of using such experimental techniques to investigate the nature of differences between "normal" and "abnormal" populations. The abnormal population used here were non-native speakers, but the basic paradigm is widely used also to analyse the language behaviour of clinical populations, and the criticism applies equally well, and probably more pressingly, to this type of work. The general problem seems to be that there is a tendency for psycholinguists to take experimental findings at a rather superficial level as "facts", rather than as effects in need of an explanation. Once they are established in this way, these facts very soon come to be adopted as tools which can be used to investigate broader theoretical positions, and to provide what looks like support for these positions. The obvious shortcoming here is that there is very little serious testing of alternative explanations of these basic effects, and given that any set of findings is always explicable in many different ways, it seems a fair bet that the current orthodoxies are likely to be mistaken. Some systematic testing of the limits of this orthodoxy therefore seems to be an urgent requirement.

These three avenues of research all have their attractions, but it has to be admitted, on the other hand, that the use of experimental

methods in this thesis has not been a wholly unqualified success. It should be apparent from these pages that the experimental tools used are far from reliable. Every one of these tools has turned out to suffer from one or more serious technical problems which had not been discussed by earlier users, and even more seriously, each of the techniques has been shown to rest on the flimsiest of theoretical foundations, which makes it difficult to find satisfactory explanations of the performance of the learner groups. In fact, one might argue that the main findings reported here have emerged in spite of the use of experimental techniques rather than because of them.

The obvious inference from this is that, given the current state of our knowledge, the questions that we have attempted to study here may be essentially unanswerable, and this is obviously a compelling reason for abandoning this particular line of approach in favour of safer paths, such as those that contain questions answerable by simpler and more easily understood more easily understood methods of research, like the observational methods we commented on in Chapter I.

To those brought up in the experimental tradition, this is obviously not an attractive notion, but what we are up against here is a clash between two rather different research traditions, the nomothetic and the hermeneutic, which have rather different priorities. Nomothetic research, concerned as it is with using controlled experiments in order to provide an explanation of the true nature of the universe through the notion of cause and effect, is obviously well able to cope with simple phenomena such as the mechanical or electro-magnetic properties of physical bodies. This type of research is much less suitable as an approach to more complex



phenomena. The hermeneutic approach, with its emphasis on understanding and interpretation seems to be much better able to cope with complex phenomena such as the behaviour of human beings to any non-trivial situation.

Ochsner (1979) has argued that in the last twenty years of second language research, a hierarchy of esteem has grown up which places experimental work above less "rigorous" types of research, and which causes hermeneutic methods such as diary studies, case reports, or interpretative speculations of the psychoanalytic type, to be correspondingly devalued. Ochsner suggests that there is very little justification for this, since the value judgements which underpin this hierarchy depend on criteria which lie outside language learning, and are thus external to the real problem. Only an a priori belief in the value of "scientific" research, for example, could give experimental methods an automatic right to be more highly esteemed than other types of approach. He goes on to argue that in the case of language learning, the nomothetic hierarchy of esteem might be topsy-turvy, and quite the reverse of what we really need. Language learning is an activity that involves the whole of a human being, and the nomothetic approach, which involves the study of isolated parts of this whole, and treats them as self-contained systems, necessarily involves abstractions and simplifications which seriously distort the overall picture. Ochsner concludes that nomothetic research in second language learning needs to be informed by a much larger body of hermeneutic research if it is to be kept in its proper perspective. This conclusion is one with which readers of this thesis will find it hard to disagree.

## APPENDICES

This section contains five sets of appendices, comprising the supplementary material for Chapters 2, 3, 5, 6 and 7. In order to facilitate reference, the sections have been numbered according to the Chapter to which they refer. There are no appendices for Chapters 1 and 4.

APPENDIX 2-A Materials used in the Experiments reported in Chapter 2.

0-order approximations:

7 words:

colère monnaie garage sac banque patience congé

10 words:

fortune réduire chaussette volontiers sorte bruyant  
sud officiel éclair votre

13 words:

attendre dent toit ignorer descendre monsieur reine  
malgré champ doux vent enfant sage

20 words:

juillet projet morceau facile dans route plutôt la  
disque appeler abattre nuage objet rester oeil humide  
dire plomb chaise pleuvoir

1st-order approximations:

10 words:

aux dont pleines nez leurs fauteuils le dans théâtre  
lumière

20 words:

avait de cette pieds la au reveil gris de femme  
pleine tête matin elle bois habitant leur que les  
toutes

2nd-order approximations:

"

10 words:

était petit mais quand tu penses profondément émue  
comme jamais

20 words:

du fromage du sel est très gentille petite fleur  
jaune et alors le perroquet\* vert foncé comme le  
premier ministre

4th-order approximations:

10 words:

disait que tu es plus gentille que moi surtout  
ennuyeuse

20 words:

aimons plus rire que quand on entend la retraite  
de vieillesse dans la plupart des jeunes gens  
élégants dont les

5th-order approximations:

10 words:

semblait drôle surtout quand il veut se reposer  
sur un

20 words:

de style moderne en dix couleurs brillantes\*\* qui  
semblaient rayonner de bonheur depuis le jour où  
sa mere avait élevé

6th-order approximations:

10 words:

essayait leur nouveau costume rose devant le grand  
ciel nuageux .

20 words:

repose à loisir sans ombre de souci heureux et  
fier encore que ce monsieur lui semble beau malgré  
son nez

notes:

\* an infrequent word; a substitution was not made here, however  
since perroquet has an obvious English cognate.

\*\* brillantes has been substituted for chatoyantes.

APPENDIX 2-B PART I

Number of words correctly recalled by each subject. 10-word strings

ORDER OF APPROXIMATION		0	1	2	4	5	6
Native Speakers:	FO1	5	5	9	10	10	10
	FO2	6	6	9	9	10	10
	FO3	3	6	10	10	10	9
	FO4	6	5	6	10	10	9
	FO5	7	7	10	10	10	10
	FO6	4	7	7	9	10	10
	FO7	8	7	6	10	10	10
	FO8	3	6	10	10	10	9
	FO9	5	6	8	10	9	9
	FIO	6	5	10	10	10	9
	FII	4	5	9	10	10	9
	FI2	4	7	8	10	10	9
	FI3	5	6	10	10	10	9
	FI4	4	7	9	9	9	9
	FI5	5	8	7	10	10	9
Learners (L <sub>1</sub> ):	LO1	6	6	5	10	4	8
	LO2	6	5	6	10	7	7
	LO3	3	5	5	9	8	8
	LO4	3	4	4	5	8	7
	LO5	3	5	2	9	6	6
	LO6	3	6	7	10	8	8
	LO7	4	6	4	4	5	3
	LO8	5	6	4	3	3	3
	LO9	3	6	6	10	9	8
	LIO	3	5	5	10	7	5
	LII	3	5	3	9	4	4
	LI2	4	5	6	7	7	3
	LI3	3	5	6	8	6	6
	LI4	3	7	6	9	6	6
	LI5	6	6	6	10	9	7
Learners (L <sub>2</sub> ):	L21	5	4	5	9	9	9
	L22	4	5	5	8	9	6
	L23	3	4	7	9	7	7
	L24	7	7	5	10	8	9
	L25	7	4	7	10	9	8
	L26	5	5	9	10	10	9
	L27	4	4	7	10	9	9
	L28	3	4	5	9	6	8
	L29	4	6	6	9	9	6
	L30	6	4	6	10	9	10
	L31	3	5	6	10	9	10
	L35	6	5	6	9	9	8

APPENDIX 2-B PART II

Number of words correctly recalled by each subject. 20-word strings

ORDER OF APPROXIMATION		0	1	2	4	5	6
Native Speakers:	F01	6	8	12	11	12	15
	F02	7	8	8	15	15	17
	F03	5	10	14	12	13	14
	F04	8	7	11	10	10	16
	F05	11	11	17	16	18	17
	F06	6	6	10	8	9	10
	F07	7	8	17	17	14	13
	F08	7	11	17	13	13	16
	F09	9	6	16	9	14	13
	F10	5	9	14	16	18	13
	F11	6	6	15	16	13	12
	F12	7	10	14	13	15	18
	F13	7	3	16	13	14	14
	F14	5	9	18	12	11	7
	F15	7	9	18	19	19	18
Learners (L <sub>1</sub> ):	L01	5	9	6	5	6	4
	L02	8	7	11	8	4	7
	L03	4	3	9	5	4	7
	L04	3	4	5	6	8	7
	L05	5	7	7	8	4	3
	L06	6	7	7	8	4	5
	L07	5	4	7	7	4	4
	L08	5	8	7	6	4	5
	L09	4	8	14	7	8	9
	L10	7	8	9	7	5	6
	L11	4	6	8	7	5	6
	L12	4	4	6	5	4	6
	L13	5	5	10	6	6	4
	L14	8	5	14	5	9	6
	L15	5	9	9	9	8	7
Learners (L <sub>2</sub> ):	L21	6	6	14	11	10	9
	L22	6	7	10	6	9	4
	L23	4	4	10	12	11	6
	L24	7	7	12	11	11	10
	L25	4	4	13	10	4	12
	L26	6	6	11	12	16	6
	L27	4	7	11	10	9	9
	L28	5	7	5	6	5	5
	L29	5	5	8	11	13	6
	L30	11	5	8	10	5	6
	L31	5	7	11	12	15	9
	L35	7	9	15	12	10	12

APPENDIX 3-A.

Materials:

Clicks occurred in each list superimposed on the underlined word.

Spanish Stimuli:

siete uno nueve dos seis cinco dies tres cuatro ocho  
tres diez cinco nueve dos siete uno ocho cuatro seis  
nueve uno cuatro cinco seis diez tres ocho siete dos  
dos seis tres diez cuatro ocho nueve uno siete cinco  
uno cuatro nueve cinco tres seis diez dos ocho siete  
nueve ocho uno siete diez tres cinco cuatro seis dos  
seis ocho siete uno cinco cuatro dos diez nueve tres  
diez siete dos tres uno nueve seis cinco cuatro ocho  
ocho cuatro tres diez cinco seis dos nueve uno siete  
cuatro dos seis ocho diez siete uno tres nueve cinco

jardín azul varios allí vino agua ahora vida llego nunca  
vamos pero cerrar negro fuego clase prisa mañana sueño ademas  
visto marchar tarde luego enero fumar camino plaza grande calma  
tierra casa digo torcer poco cesta alto año monte llover  
guerra llamar largo salir entonces ruido novia linea mano cuarto  
hacia flores suelo triste rama pequeño nada bajo contar tanto  
pidió había gota hermano haber árbol nueva sacar pueblo dinero  
gente todo cama fino estar sobre jefe ventana cuando techo  
llevar delgado tiempo padre entre pino durante fondo calle verano  
señora derecho comer abrir sentada llave desde primera puente sierra  
había una vez una familia de ratones que vivían en una casa  
el ama de la casa había comprado un gato feroz  
y cuando los ratones salían al anochecer en busca de comida  
el gato les seguía y trataba de matarles  
los ratones se dijeron que si el gato tuviera un cascabel  
alrededor del cuello  
no podría esconderse detras de nada para atacarles de noche  
y un día mientras el gato estaba durmiendo en un rincón  
todos los ratones se acercaron a él para ponerle el cascabel  
pero el ruido que hicieron despertó al gato que se los comió a todos.

## English Stimuli

seven one nine two six five ten three eight four  
three ten five nine two seven one eight four six  
nine one four six five ten three eight seven two  
two six three ten four eight nine one seven five  
one four nine five three six ten two eight seven  
nine eight one seven ten three five four six two  
six eight seven one five four two ten nine three  
ten seven two three one nine six five four eight  
eight two six four nine one seven five three ten  
four five eight six three nine two ten seven one

distant lesson surely secret balance habit praise credit thick sheep  
brush spent fought avoid engine maybe level print useful asleep  
steam china police thrown orange capital bought source basket slave  
detail clothes silent render largely profit correct wander inquire merely  
blind search admire grace honest pattern meaning unite cream nurse  
private scatter central cease arrange angry deliver distant factory  
mistake

citizen climb beach flour owner defence brain steel region partly  
slept proud honey castle design metal attack collect bitter couple  
sunday empty throw insist string afford depend funny royal loose  
message spoil sheet fortune route handle theatre fence compare nobody

there was once a young rat named Arthur  
who could never make up his mind  
whenever his friends invited him to go out with them  
he couldn't think whether he ought to go or stay at home  
and on the day when all the other rats moved to a new house  
where they thought they'd be able to find more food  
Arthur couldn't bring himself to a decision of any kind  
so he stayed behind until eventually he starved.



APPENDIX 3-B.

Percentage of clicks corerectly located.

Type of Material:		SD	SW	SS	ED	EW	SS
Subject Number:	S01	20	00	00	10	40	25
	S02	80	80	00	90	100	25
	S03	60	100	77	80	100	50
	S04	90	100	22	90	90	12
	S05	100	100	62	100	100	25
	S06	70	80	66	100	100	00
	S07	70	90	44	90	100	25
	S08	80	100	77	100	100	25
	S09	40	10	00	10	20	00
	S10	44	90	44	90	100	57
	EO1	40	80	22	30	90	25
	EO2	100	100	89	100	100	75
	EO3	20	60	00	50	90	25
	EO4	00	20	00	20	100	25
	EO5	40	40	38	90	70	50
	EO6	20	20	00	10	90	25
	EO7	00	00	11	10	100	12
	EO8	00	30	33	30	40	00
	EO9	20	00	00	10	40	11
	E10	20	40	00	50	100	38

APPENDIX 3-C.

Basic Displacement Scores.

In this score, all displacements are counted as positive. One point is given for each position that the click is moved from its objective position, where position is defined as a word or the space between two words.

The total displacement score is then divided by the number of sentences in order to reach the figures presented below.

Stimulus Material:		SD	SW	SS	ED	EW	ES
Subject Number;	SOI	1.2	1.2	2.8	1.6	1.2	2.5
	S02	0.3	0.2	1.9	0.2	0.0	1.9
	S03	0.5	0.0	0.3	0.4	0.0	1.0
	S04	0.3	0.2	1.5	0.2	0.0	2.12
	S05	0.2	0.0	0.75	0.0	0.2	2.0
	S06	0.6	0.4	0.66	0.0	0.0	2.75
	S07	0.6	0.2	1.3	0.2	0.0	2.0
	S08	0.4	0.0	0.4	0.0	0.0	2.25
	S09	1.1	1.8	2.0	1.8	1.6	2.9
	S10	1.2	0.2	1.3	0.2	0.0	0.75
	EOI	1.9	0.3	2.3	0.7	0.1	2.0
	E02	0.0	0.0	0.22	0.0	0.0	0.5
	E03	0.9	0.4	1.3	0.6	0.2	1.38
	E04	1.3	1.3	3.1	1.1	0.0	2.0
	E05	1.1	1.1	1.5	0.4	0.4	1.25
	E06	1.0	1.2	2.6	1.1	0.1	1.38
	E07	1.8	0.9	1.2	1.4	0.0	3.25
	E08	1.8	0.9	1.2	1.4	0.0	3.75
	E09	0.8	1.0	2.8	0.9	0.6	2.25
	E10	0.8	0.6	1.3	0.8	0.1	1.1

APPENDIX 3-D.

Corrected Displacement Scores.

These scores are calculated by working out the total displacement of the clicks for each subject, and dividing this total by the number of strings in which clicks were displaced.

Stimulus Material:		SD	SW	SS	ED	EW	ES
Subject Number;	SOI	1.5	1.2	2.8	1.8	2.0	3.3
	S02	1.5	1.0	1.9	2.0	0.0	2.5
	S03	1.25	0.0	1.5	2.0	0.0	2.0
	S04	1.5	2.0	2.0	2.0	0.0	2.4
	S05	2.0	0.0	2.0	0.0	2.0	2.6
	S06	2.0	2.0	2.0	0.0	0.0	2.75
	S07	2.0	2.0	2.4	2.0	0.0	2.3
	S08	2.0	0.0	2.0	0.0	0.0	2.7
	S09	1.8	2.0	2.0	2.0	2.0	2.75
	S10	2.4	2.0	2.4	2.0	0.0	2.0
	EOI	3.2	1.5	3.0	1.0	1.0	2.7
	E02	0.0	0.0	2.0	0.0	0.0	2.0
	E03	1.13	1.0	1.33	1.2	1.0	1.83
	E04	1.4	1.6	3.1	1.6	0.0	2.7
	E05	1.38	1.8	2.3	1.3	1.3	2.5
	E06	1.25	1.5	2.6	1.2	1.0	1.8
	E07	1.8	1.0	1.6	1.6	0.0	3.3
	E08	1.4	2.3	1.0	1.6	2.3	2.75
	E09	1.0	1.0	2.9	1.0	1.0	2.25
	E10	1.14	1.0	1.33	1.6	1.0	1.8

APPENDIX 5-A.

Materials used in the Pilot Study.

A complete list of the materials used is given below. This list also shows the order of presentation, which was the same for all subjects.

KEY: S: Spanish; E: English; V: Verb; N: Noun; A: Adjective

1. comer queso	SVN	2. earth glass	ENN
3. alone heavy	EAA	4. beber conac	SVN
5. other early	EAA	6. enjoy music	EVN
7. bolso vacío	SNA	8. reina calor	SNN
9. fresh fruit	EAN	10. prisa nieve	SNN
11. tomar parte	SVN	12. mujer guapa	SNA
13. fuego sitio	SNN	14. rubio fácil	SAA
15. queen month	ENN	16. happy child	EAN
17. tarde feliz	SAA	18. lápiz negro	SNA
19. teach class	EVN	20. large ready	EAA
21. green grass	EAN	22. sucio gordo	SAA
23. first whole	EAA	24. spend money	BVN
25. triste claro	SAA	26. empty house	EAN
27. scene blood	ENN	28. crowd paper	ENN
29. playa reloj	SNN	30. viaje largo	SNA
31. tener miedo	SVN	32. drink water	EVN

APPENDIX 5-B.

Pilot Study: Scores of Individual Subjects.

A. TOTAL WORD SCORE:

LANGUAGE PAIR TYPE		ENGLISH				SPANISH			
		NN	AA	AN	VN	NN	AA	NA	VN
Subject	G01	8	7	8	8	4	5	6	7
	G02	8	8	8	8	7	7	5	7
	G03	7	6	8	8	6	6	4	5
	G04	8	5	7	8	5	5	6	6
	G05	7	6	7	8	5	6	6	7
	G06	6	8	8	8	6	5	5	7
	G07	7	8	8	8	7	7	7	7
	G08	5	5	8	8	4	3	8	7
	G09	6	8	8	8	4	7	7	5

B: WHOLE PAIR SCORES:

LANGUAGE PAIR TYPE		ENGLISH				SPANISH			
		NN	AA	AN	VN	NN	AA	NA	VN
Subject	G01	4	3	4	4	1	1	2	3
	G02	4	4	4	4	3	3	2	3
	G03	3	2	4	4	2	2	0	1
	G04	4	1	3	4	1	1	2	2
	G05	3	2	3	4	2	1	2	3
	G06	2	4	4	4	2	1	2	3
	G07	3	4	4	4	3	3	3	3
	G08	2	1	4	4	1	1	4	3
	G09	2	4	4	4	1	3	3	3

APPENDIX 5-C.

Materials used in the Main Experiment:

A complete list of the materials used is given below.

KEY: S: Spanish; E: English; A: Adjective; N: Noun

ENGLISH STIMULI:

brief visit	EAN	loose awful	EAA
white paper	EAN	smart still	EAA
black cloud	EAN	brown heavy	EAA
false teeth	EAN	alive equal	EAA
fresh fruit	EAN	noble alone	EAA
funny story	EAN	brave apart	EAA
sharp knife	EAN	royal quick	EAA
angry crowd	EAN	rough great	EAA
quiet voice	EAN	grand sweet	EAA
green grass	EAN	blind clean	EAA

SPANISH STIMULI:

bolso vacío	SNA	santo firme	SAA
perro feroz	SNA	sucio feliz	SAA
viaje largo	SNA	capaz manso	SAA
gente seria	SNA	lleno breve	SAA
golpe flaco	SNA	rubio fácil	SAA
lápiz negro	SNA	ancho listo	SAA
campo verde	SNA	corto mejor	SAA
aldea pobre	SNA	sagaz claro	SAA
coche viejo	SNA	*suave chico	SAA
mujer guapa	SNA	tarde gorde	SAA

\* This sequence was deemed to be ambiguous, and was not included in calculating the mean recognition thresholds.

APPENDIX 5-D.

Main Experiment: Scores of the Individual Subjects.  
Mean recognition thresholds in msec.

LANGUAGE PAIR TYPES		ENGLISH		SPANISH	
		AA	AN	AA	NA
GROUP B	B01	60.0	60.0	73.3	68.9
	B02	70.0	68.0	71.6	72.5
	B03	51.0	49.0	53.8	75.5
	B04	55.0	53.0	70.0	72.5
	B05	64.0	70.0	78.1	86.1
	B06	65.0	54.0	80.0	83.1
	B07	70.0	60.0	77.8	81.0
	B08	69.0	65.0	78.8	78.3
	B09	67.0	61.0	73.3	72.2
	B10	62.0	58.0	78.8	82.2
GROUP A	A01	64.0	51.0	76.6	74.5
	A02	63.0	61.0	78.9	71.0
	A03	62.0	63.0	68.8	78.0
	A04	55.0	57.0	60.0	63.0
	A05	65.0	65.0	75.5	73.0
	A06	63.0	58.0	63.3	74.0
	A07	67.0	59.0	65.5	68.0
	A08	51.0	48.0	58.8	64.0
	A09	69.0	64.0	74.4	70.0
	A10	72.0	66.0	81.6	77.0
GROUP S	S01	84.0	81.0	102.2	81.0
	S02	87.2	90.0	90.0	93.5
	S03	77.5	78.0	67.8	73.3
	S04	82.5	68.0	83.3	71.0
	S05	108.5	115.0	104.4	102.5
	S06	87.5	94.5	82.2	82.5
	S07	113.9	108.3	108.3	82.0
	S08	74.0	69.0	80.0	64.0
	S09	75.0	77.5	75.0	67.0
	S10	77.0	80.5	84.4	77.0

APPENDIX 6-A-1.

Experiment 6-1 Materials.

Test sentences in order of presentation. Sentences not preceded by any sign are miscellaneous. Sentences preceded by (a) represent structure (a), those preceded by (b) represent sentence type (b). Figures following this indicator show the position at which the sentence token was probed. Response words were chosen so that uncertainties due to inaccurate timing with certain low amplitude sounds were avoided.

1. los niños durmieron bien anoche
2. la casa parecía inocupada
3. mi amigo nació en Barcelona
4. el coche seguía la carretera de Madrid
5. los estudiantes no tenían dinero
6. los pájaros cantaban en los árboles
7. a2 el negociante rico prefiere coches franceses
8. mi madre acaba de volver del mercado
9. a1 la muchacha tonta comió manzanas verdes
10. el médico visita los pueblos pequeños también
11. b2 la secretaria escribió cuatro cartas importantes
12. a4 el estudiante listo lee libros serios
13. a3 la mujer hermosa llevaba zapatos negros
14. b3 el oficial sacó varios billetes azules
15. b1 el profesor compró varios libros raros
16. la casa estaba en lo alto de la ciudad
17. b4 el turista dejó dos maletas pesadas
18. mañana tenemos que escoger un nuevo lugar para el mítin
19. a1 el rey sabio decreta leyes justas
20. b3 el ladrón voló tres pinturas modernas
21. b2 el obispo abrió siete iglesias nuevas
22. a2 el arquitecto alemán dibujó edificios enormes
23. este maletín contiene varios papeles importantes
24. a3 el viajero inglés frecuentaba países lejanos
25. b4 el guardia paró algunos autocares vacíos



26. tengo que estar de vuelta en la ciudad a las ocho de la mañana
27. a4 mi hermano mayor compró pantalones verdes
28. b1 el criado cerró varias puertas enormes
29. b1 mi abuela tenía ocho gatos negros
30. b3 la niña pidió unas peras maduras
31. las maletas no caben en el coche
32. a1 el bandido cruel cometió delitos violentos
33. mejor será que pasemos por la comisaría
34. a3 el niño enfermo quería juguetes nuevos
35. a4 el empleado perezoso evitaba tareas difíciles
36. b4 el artista pintó cuatro ollas blancas
37. tiene mucha sed y quiere tomar algo antes de salir
38. a2 el joven inteligente coleccionaba sellos extranjeros
39. b2 el pastor buscaba veinte ovejas perdidas
40. b4 el actor fumó muchos cigarrillos baratos
41. los obreros iban a trabajar en la fábrica
42. a3 la chica delgada remendaba faldas viejas
43. b3 el alumno resolvió varios problemas difíciles
44. el abogado llevaba un traje gris y una corbata negra
45. a4 el presidente nuevo concedió libertades políticas
46. a1 el soldado valiente recibió heridas mortales
47. este remedio siempre da buen resultado
48. b1 el campesino quería dos caballos fuertes
49. a2 el director avaro buscaba provechos enormes
50. b2 el obrero bebió tres cervezas frescas

APPENDIX 6-A-2.

Experiment 6-1.

Median response latency in msec to four probe positions in two Spanish sentence types; native speakers and learners of Spanish.

SENTENCE TYPE	A				B			
	1	2	3	4	1	2	3	4
NATIVE SPEAKERS:								
SOI	750	810	780	880	960	860	760	820
SO2	1530	1593	1377	1562	1588	1602	1530	1260
SO3	1120	1046	780	800	800	1000	830	810
SO4	1140	1100	900	1000	1320	1140	1120	845
SO5	780	970	700	700	780	625	790	700
SO6	730	1000	740	840	763	820	750	800
SO8	770	940	843	740	800	780	680	720
SO9	1020	1190	1410	1030	1080	1010	1120	1083
SIO	1300	2100	1160	1220	1160	1660	1050	1190
SII	800	820	1000	963	980	1020	918	960
SI2	1960	1600	1040	1140	1200	1450	1160	1290
LEARNERS:								
LOI	1695	1508	1260	1140	1403	1020	840	1005
LO2	905	1160	800	1100	1160	670	900	1120
LO3	1140	1240	1530	1000	950	1000	950	980
LO4	660	1100	930	640	960	880	1000	820
LO5	1140	1330	1070	1060	1010	1040	970	855

APPENDIX 6-B-1

Materials used in Experiment 2.

Test sentences are shown in order of presentation. Sentences not preceded by any distinguishing mark are miscellaneous. Sentences preceded by (a) represent structure (a); those preceded by (b) represent structure (b). Figures following this indicator show the position at which each sentence was probed.

1. los niños durmieron bien anoche
2. la casa parecía desocupada
3. mi amigo nació en Barcelona
4. el coche seguía la carretera de Madrid
5. los estudiantes no tenían dinero
6. los pajaros cantaban en los arboles
7. b2 el joven inteligente coleccionaba mariposas
8. b1 el criado gordo bebía vino
9. mi madre acaba de volver del mercado
10. a3 el gobierno concedió libertades politicas
11. a1 mi abuela fumaba cigarrillos baratos
12. b3 el nino enfermo quería juguetes
13. la casa estaba en lo alto de la ciudad
14. a2 el director buscaba provechos inalcanzables
15. b1 el rey sabio escribió libros
16. tenemos que escoger un nuevo lugar para el mítin
17. este maletin contiene muchos papeles importantes
18. a3 el empleado evitaba tareas difíciles
19. a1 el campesino quería caballos fuertes
20. tengo que estar de vuelta a las ocho de la mañana
21. b3 la chica delgada remendaba faldas
22. b2 el negociante rico bebía coñac
23. las maletas no caben en el coche
24. a2 el viajero frecuentaba paises lejanos
25. a3 mi hermano compró pantalones verdes

26. b2 el artista alemán pintaba iglesias
27.     es mejor que vuelva mañana
28. a1 el profesor compraba libros raros
29. a2 la mujer llevaba zapatos negros
30.     tiene mucha sed y quiere tomar algo antes de salir
31. b1 la muchacha tonta comió demasiado
32. b3 el obrero cansado bebió cerveza
33.     el abogado llevaba un traje gris y una corbata negra
34. a2 el autor escribió poemas largos
35. b2 el tendero avaro contaba billetes
36. b1 el soldado valiente recibió heridas
37.     este remedio siempre da buen resultado
38. a1 el bandido cometía delitos violentos
39. a3 el estudiante leía libros serios
40. b3 el actor calvo buscaba trabajo

APPENDIX 6-B-2

Median response latency in msec to three probe positions in two Spanish sentence types.

SENTENCE TYPE	A			B		
	1	2	3	1	2	3
NATIVE SPEAKERS:						
S01	1045	1030	965	995	1000	1015
S02	860	990	940	935	995	835
S03	1075	1125	1150	1120	1190	1120
S04	1310	1765	1550	1240	1320	1420
S05	1045	1205	1025	1180	1145	1040
S06	1495	915	1070	1320	1140	900
S07	815	735	740	700	735	830
S08	865	785	850	855	805	830
S09	1070	935	755	960	955	865
S10	870	1085	900	875	950	940
S11	820	1230	970	1015	1225	940
LEARNERS:						
L01	1225	1100	955	1260	1360	940
L02	1495	1235	850	1370	1630	1800
L03	1160	1285	1220	1230	1440	1040
L04	1810	970	1090	870	1120	1090
L05	990	990	1110	940	1110	1000
L06	1305	1165	930	1160	1340	1120
L07	1265	1290	1165	1200	1450	1205
L08	995	1450	1115	750	1670	975
L09	1180	1180	995	1040	1570	955
L10	1160	1125	1220	1230	1440	1040
L11	1160	1210	1440	1302	1365	1350
L12	1100	1180	1080	1195	1310	2050
L13	1260	1305	1060	1270	1375	1070
L14	1090	960	805	965	950	1090
L15	995	1090	920	920	1095	905
L16	865	900	780	920	870	760

### APPENDIX 6-B-3.

#### Error rates.

In addition to the major differences between the groups reported in Section 6-6-5, there was a marked difference in error rates between the two groups. Errors produced by native speakers were minimal, while the learners produced a large number of incorrect responses. The error rate of the native speakers was 2.5%, while that of the learners was 22%. Errors do not appear to be related to Sentence Type or differentiated according to Probe Position, however. The learners also frequently failed to respond at all to a probe word; only one such failure was recorded for the native speaker group.

These findings suggest that even with the very short sentences used in this experiment, the learners were experiencing considerable difficulty. The figures, impressive though they are, actually mask the full extent of the problem. Some learners tested (N=5) failed to produce any scores at all for one or more of the possible probe positions. These subjects were discarded, and their scores were not included in the main analysis or in the error rate figures quoted above. The summary figures for error-rates also exclude instances of non-responding where the answer was so hesitant that it failed to trigger the voice key. Both these factors would have served to increase the already large difference between the groups. It is possible that the high error rate of the learners might be due to the use of infrequent words which were outside the range of the subjects' recognition vocabulary. It did not prove possible

to construct the requisite number of sentences from a vocabulary limited only to the most frequent items. Care was taken, however, to ensure that all the probe words and correct response words fell within the expected range of words known by the learner group. Furthermore, when asked about their inability to produce a response, most of the subjects commented that they had understood the sentence perfectly well, but that they were simply unable to remember the response word..

Reports of this kind are not wholly reliable, of course, but they perhaps suggest that one of the main differences between understanding a sentence in a foreign language and ones native language is that in the former case the record of the input is lost much more rapidly than is the case for native language input. This suggestion clearly deserves further study.

## APPENDIX 7-A.

### The Language of the Experiments.

Three of the four sets of experiments reported in the preceding chapters deal with the performance of learners of Spanish, while the fourth examines the performance of learners of French. At the time this work was carried out, this inconsistency was not considered important. Spanish and French are both Romance languages, and have many syntactic and morphological features in common, as well as a large common lexicon, and these features differentiate both languages from English. The choice of French for the experiments in Chapter 2 was dictated principally by practical considerations, such as the availability of large numbers of native English speakers learning French at various levels, and the existence of a published set of statistical approximations to French that had already been used in experimental work. Native speakers of English learning French were studied in preference to native speakers of French learning English because it was felt that prolonged residence abroad might be a factor that produced changes in structuring behaviour. All the available native speakers of French would have lived in Britain for some time, and thus received heavy exposure to English. It was assumed that, within these limits, the language tested was of secondary importance, and that all learners of whatever language would show behaviour that was essentially comparable. With hindsight, this assumption appears to be naive. In fact, it is a clear case of the universalist assumptions that were criticised earlier in this chapter. However, if we criticise this assumption here, we must



inevitably raise the question of whether the differences found between native speakers and learners of French would be likely to repeat themselves in the case of Spanish, or whether the Statistical Approximation effect is a language specific one, as we have suggested might be the case for some other structural effects in English.

French and English resemble each other in that both have a strict normal SVO word order, and in this respect they differ from Spanish, where, as we have seen, word order is much more flexible. There are some grounds for believing that Spanish approximations, particularly higher orders of approximation, might be more difficult to handle than their English or French counterparts on account of this characteristic.

The main feature that differentiated higher order approximations from fully coherent text is that in the former some whole constituents play an ambiguous role, acting as parts of two incompatible larger constituents at the same time. Examples of this type of ambiguity can be found in the following passage taken from Miller and Selfridge (1950):

"5th-order approximation to English: (40 words)

working towards a goal for his team in old New York was a wonderful place wasn't it even pleasant to talk about and laugh hard when he tells lies he should not tell me the reason why you are is evident .....

Ambiguous constituents in this passage are:

Old New York	(Object of <u>in</u> and subject of <u>was</u> )
wasn't it	(tag phrase to <u>wonderful place</u> and Main Verb of the following phrase.)
when he tells lies	(subordinate to both <u>laugh</u> and <u>tell</u> )
the reason why you are	(object of <u>tell</u> and subject of <u>is</u> )

Between these ambiguous phrases, the text is fairly coherent, however. Now the relatively fluid order of constituents in Spanish compared with French and English, and the fact that verbs in Spanish are frequently to be found without overt subjects means that the possibility of ambivalent phrases occurring in approximations to Spanish is much higher than in English or French. This means that the very long sequences of coherent words that characterize high orders of approximation in English and French would be less likely to materialize. It seems, therefore, likely that high orders of approximation to Spanish should be less easy to handle than corresponding orders of approximation in these other languages.

There are no published sets of approximations to Spanish, and no published norms for the behaviour of native speakers of Spanish on this type of stimuli. However, Table 7-A-1 below shows a set of approximations to Spanish collected from a group of native Spanish speakers temporarily resident in Britain. The method used for collecting these was the same as that described by Miller and Selfridge (1950). There are, as predicted, a number of examples of ambivalent constituents in the higher orders, particularly in the 4th-order approximation, where the verb se murió appears to have two subjects, one preceding it and the other following it. This configuration would be impossible in English of course, or in French. Given the shortness of the strings in this sample, ambivalent constituents of this type do not appear sufficiently frequently for it to be likely that they would produce very marked differences in native speaker behaviour. This claim is supported by the evidence shown in Figure 7-A-1. This data shows the performance of a small group of native Spanish speaking subjects on a set of 20-word strings

taken from the approximations in Table 7-A-1. This data differs slightly from the English and French data discussed earlier, in that the rapid increase in recall levels usually found on 2nd- and 3rd-order strings of this length is missing, and instead, the data show a gradual increase in performance from 0-5th order, a pattern more characteristic of longer strings of words. Though the 5th-order approximations to Spanish produce scores which are comparable with 5th-order French and English scores, the clear asymptote usually found with 3rd, 4th and higher order approximations is missing here. This result is in line with a prediction made in Meara (1975) where it was argued that the severe short term constraints that operate in Spanish - e.g. Noun Phrase Concord, - would prevent the build up of long meaningful sequences in Spanish approximations, and thus render middle order approximations to Spanish rather more difficult than their English counterparts.

A very similar set of data can be found in Figure 7-A-1a which shows the results of a group of native Welsh speakers on a set of approximations to Welsh drawn from the lists in Table A-7-2. Welsh, a language which differs markedly from the other languages discussed in this section, ought to produce approximations which are very difficult to handle because of its characteristic VSO word order. The results do indeed suggest that middle orders of approximation are rather hard. As was the case for Spanish, the recall scores rise gradually, and the asymptote is missing, but the scores on 5th-order Welsh are again broadly comparable with those of English and French, and both sets of scores rise more steeply than is the case with the learners of French reported in Chapter 2.

With some reservations, then, about middle orders of approximation, it seems to be the case that immediate recall of statistical approximations is a robust task, which produces broadly similar results in a wide range of very different languages. In all the cases reported increased contextual constraint leads to better recall, though there is some variation in the slope of this performance curve. This seems to be sufficient justification for the generalizations made earlier. It should be borne in mind, however, that immediate recall of texts, where only the number of words correctly reported are scored, is a very crude measure of the differences between texts, and it is very likely that a detailed analysis of approximations to various languages would show up systematic differences that might be expected to have behavioural consequences which could be measured with more sensitive experimental tools.

Table 7-A+1

Statistical approximation to Spanish:

2nd-order approximation:

cuando llegue cuando yo tengo siete enanos viven en pantalones  
te pronto tu camisa nueva que viene de madera está pensando  
todo Madrid es hermoso pasear bajo que me gusta mucho y pan  
de vosotros al cine de Italia algún día

3rd-order approximation:

mientras las flores estaban secas las hojas caen en el jardín  
de María está comiendo carne de vaca fresca es mejor ésta  
que yo soy ésa que elegiste es horrible sentir calor al  
acercarme a la francesa no quiere venir pero no

4th-order approximation:

cuando mi tío se murió el pájaro de María está muerto de frío  
eres antipático cuando te ríes me gustas mucho más si tú  
quieres venir al centro mañana por la mañana temprano me  
levantaré si me despierto pronto en el

5th-order approximation:

tenía que pedir a mi padre permiso para salir al pueblo hay  
que seguir ésta sesión mañana por la tarde iré al cine como te dije  
ayer lo mismo que hoy

6th-order approximation:

encontró a la niña que había visto el otro coche acelerar  
rápidamente pero no me fue posible porque el trabajo es  
lo que dignifica al hombre es la libertad mientras lo contrario  
no sea de gran importancia no me molesta

Table 7-A-2.

Statistical approximations to Welsh.

2nd-order approximation:

drannoeth daeth y mae'r wyf yn y byd er garwed y gwanwyn gwelais  
olau llachar heddiw yw'r arian wedi yfory os byddaf yn canu  
yn mochyn adref wedi mynd 31 words

3rd-order approximation:

tybiodd y dylwn fynd i'r ysgol mae un o'r mynydd a choedwig las  
yn un hefyd o flaen y 21 words

4th-order approximation:

caerddodd drwy gydol y flwyddyn y bu yma gwellodd lawer o welliannau  
am ddyfodol gwych i'r hen yn un trwy'r ddau steision cawsom luniaeth  
a sgwrs gan y 30 words

5th-order approximation:

aeth y frenhines ati i ofyn os oedd hi yn well o lawer ar Gymru y  
gwyliau yma bydd Mona yn byw mewn carafannau ar y traeth mae  
plant yn brysur wrth y ddesg i geisio 36 words

6th-order approximation:

Maen nhw wedi eu gwerthu yn y farchnad aeth y llestri yn rhad ond  
gwelais nad yw hyn yn ddignonol i'ch cadw rhag oeri bwyteuach ddigon  
dyna blant lwcus ydych chi yn cael anrheg 35 words

7th-order approximation:

gofynnodd yr un cwestiwn ym mhob man yr aeth iddo gwelodd olion  
ei gyd-filwyr ac o'r tywyllwch mawr clywodd sgrech uchel a rhedodd  
nerth ei phen am gynorthwy at 30 words

8th-order approximation:

cyn gynted ag y daethant o fewn golwg y ddinas gwelasant adar y  
nefoedd yn farw ac aethant yn brin eu cynnyrch ar ôl y streic ond  
gwelodd mai parod oedd y dynion i'w gredu ef er 38 words

Figure 7 A-1.

Immediate recall of  
Statistical Approximations  
to Spanish and Welsh

Figure 7 A-1a

Spanish Data

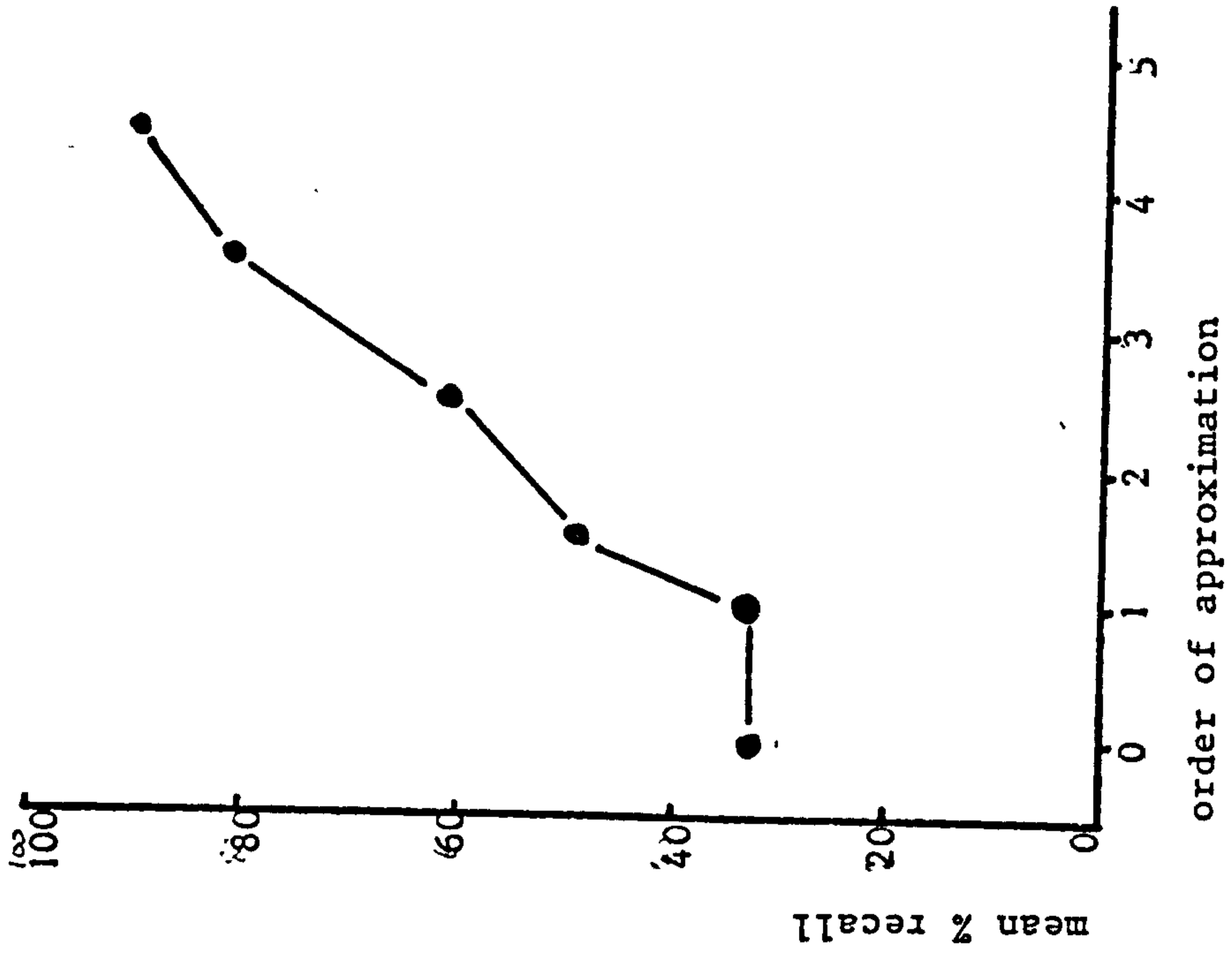
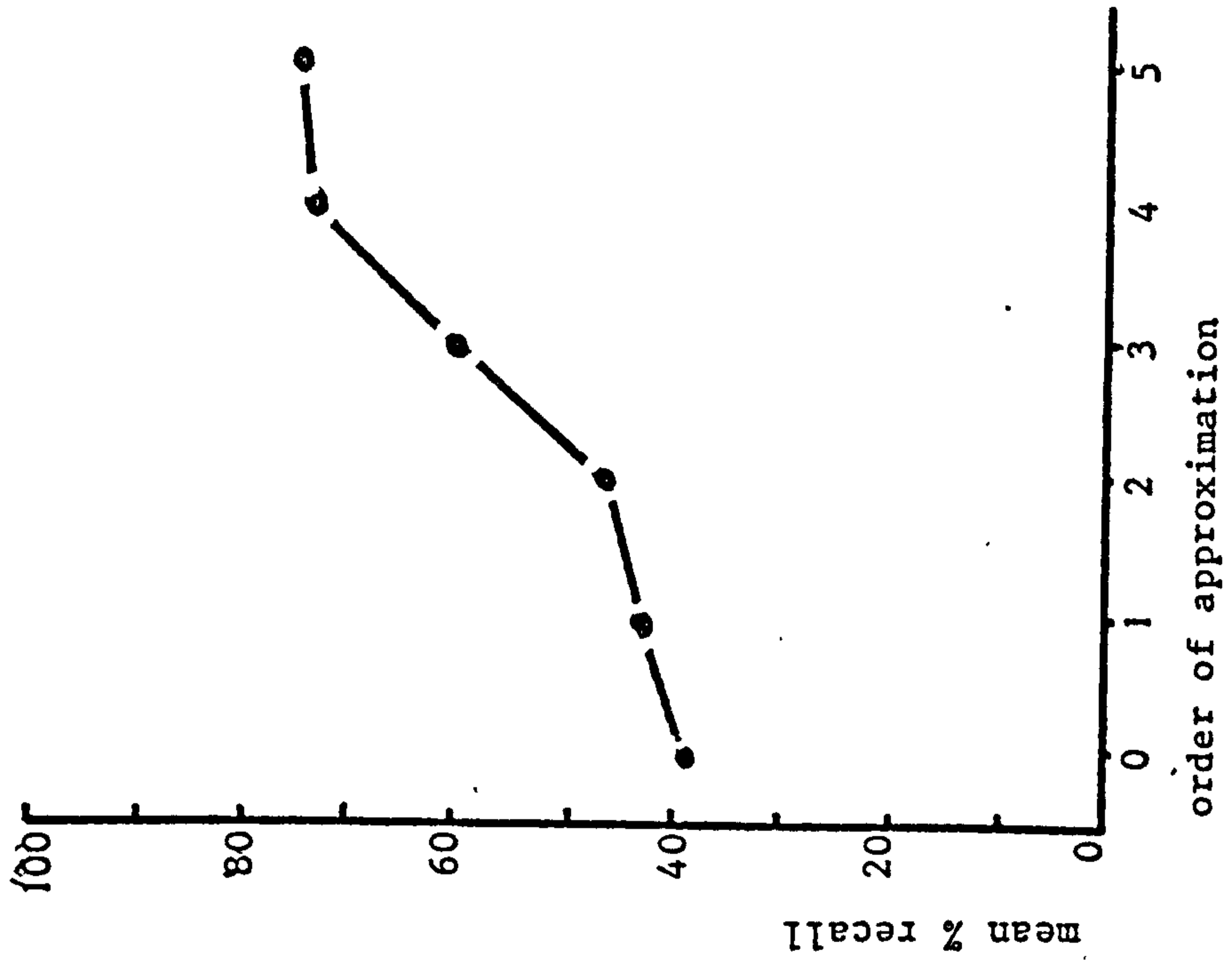


Figure 7 A-1b

Welsh Data



## APPENDIX 7-B.

### The Level of Proficiency of the Subjects.

One major criticism that might be levelled at the experimental work reported in the previous chapters is that there is no independent measure of the level of proficiency of the learner subjects in their target language. The subjects were classified according to the level of course that they were following (e.g. A-level, 1st-year degree course, etc.) or according to examinations that they had already passed (e.g. graduates in Spanish.). In doing this we have been following what is normal practice in experimental work of this type - cf. Albert and Obler (1979) *passim*. It would obviously have been interesting to obtain independent measures of the subjects language ability and to relate these to performance on the experimental tasks. This was not done, however, for both practical and theoretical reasons

The main practical objection which influenced this decision was that there are no standard tests of ability in French or Spanish that could have been used conveniently. A number of test batteries do exist, but these tend in their very nature to be large and cumbersome. The Cooperative Tests produced by the Modern Language Association of America are sometimes used for experimental purposes, for example, but these tests take over an hour to administer, and this time factor effectively rules out these tests as a practical proposition unless one has large financial resources or a set of captive subjects willing to spend several hours on an experiment.



One alternative to a test battery would have been to use a test of the type described by Lambert (1955). In this paper, Lambert describes a simple motor task (responding to commands in two languages) and in a later paper, (Lambert et al. (1959)) he showed that the scores of language dominance based on the results of this test correlate reasonably well with a range of other tests in English and French. Lambert's tests were not used here, however. It was felt that the motor task was not a fair assessment of subjects' ability in the second language, as it relies exclusively on ability to handle colour words, numbers and the words for right and left, and thus lacks the necessary high level of face validity. Lambert's other tasks are also rather crude, often difficult to adapt for languages other than French - his "hidden words" task, for example where the Subject is required to recognize words in English and French in letter sequences such as DANSODENT is one which the characteristic distribution of word lengths in English and Spanish would make almost impossible. Furthermore, though Lambert claims that his tests show a high degree of intercorrelation, none of the tests has ever been used in conjunction with larger, well-tried test batteries, and so the relationship between Lambert's work and normal expectations of learner's performance in a second language is unknown.

The only viable alternative way of testing the learner subjects was to use Cloze Tests, which though they are not standardized are easy to construct, simple and quick to administer and easy to score.

Cloze tests were first designed by Taylor (1953) as an alternative to the then existing formulae for assessing the readability of a text. Instead of counting objective features of the passage to be assessed - features such as mean sentence length, mean word

length, mean word frequency, and so forth, which are then weighted according to complex formulae such as that of Flesch (1948) - Taylor deleted words from the passage in a regular fashion - say, every seventh word - and asked a large group of native speakers to fill in the blanks in the mutilated passage. Words correctly restored scored one point. Taylor argued that a passage that produces a large number of correct responses will be one that is easier to read than one that produces a much smaller number.

This basic idea has been adapted by a number of authors, notably by Oller (1973) and elsewhere into a tool for assessing the ability of second language learners in their target language. Oller argues that as well as its use in assessing texts, Cloze test procedure can also be used to assess the relative skills of individual speakers. Speakers who produce a large number of correct responses for a given text might be assumed to have understood the text better than speakers who produce a lower number of correct responses, and that comprehension ability measured in this way might reasonably be expected to correlate with other language skills. Oller went on to show that Cloze Tests used in this way clearly distinguished between native speakers and even advanced learners of a language. In an important paper (Oller and Conrad (1971)) he further showed that the scores of a group of learners of English on a series of Cloze tests correlated highly with their scores on a large standard battery of Proficiency tests, and on the strength of this finding he argued that Cloze tests might be used in place of Proficiency batteries. In a later experiment, (Oller et al. (1972)) Oller showed that more lenient scoring methods such as counting as correct not just the original word, but any acceptable alternative did not significantly

reduce the size of this correlation. Oller concludes that Cloze tests are a reliable and robust measure of overall ability in the foreign language.

Although Oller's main concern is with testing proficiency, it is clear that, in principle, at least, Cloze tests are likely to be a tool of some importance for experimental psycholinguists. However, a good deal of background work will be necessary before they can be used with any confidence in experimental research.

Cloze tests were used with some of the experiments reported in the preceding chapters, as a way of controlling the subjects' ability in their second language. In particular, all the subjects tested in the Chapter 6 (Probe Latencies) were required to fill in a short Cloze Test which is reproduced below, in Table 7 B-1. All subjects scored at least 50% on this test using the exact word criterion. A similar test was used with the subjects studied in Chapter 3, and similar results obtained. It was felt, however, that the Cloze Test did no more than corroborate impressionistic judgements made by the experimenter. Since all the experiments involved individual sessions run by the experimenter, judgements of this sort were relatively easy to make, and the additional information supplied by the Cloze Tests was not particularly illuminating. For the experiments used in Chapters 5 and 6, a rather different type of control was devised, and it was generally felt that this was more successful as a way of weeding out subjects whose ability was very low. In these experiments scores were taken only when the correct answer was produced, and where subjects failed to produce the correct answer either through error or through ignorance, their scores do not contribute to the group mean score. An inbuilt control of this type clearly has much

to recommend it, as there can be no doubt that the subjects did understand the material on which they were tested.

There are, in any case, two major problems with Cloze tests that seem to have attracted very little attention, but which suggest that some caution is needed in accepting at face value all the claims that are made on their behalf.

Firstly, almost all the work that has been done on Cloze tests has been carried out in English - just like the greater part of psycholinguistic work in general - and though there is a growing body of work on the South East Asian group of languages (e.g. Oller et al (1972) or Anderson (1976)) the crucial test that correlated Cloze Scores and overall ability in a foreign language has only been carried out in English. It may seem reasonable to assume that the same level of correlation would be found with any language, but this is not necessarily the case, and there are in fact several reasons for believing that Cloze tests in Spanish ought to be relatively more difficult than a passage based on similar material in English - say, for example, a Spanish and an English translation of a test in a third language.

Spanish texts are generally shorter than their English equivalents, and this means that each deletion in a Spanish Cloze Test contains more information than in an English one. Furthermore, Spanish passages typically have higher Type-Token Ratios than English passages of equivalent length, which suggests that Spanish typically uses a wider range of vocabulary than English does. Both these factors can be seen at work in the sentences below;

a) every night she used to go to bed at midnight

b) cada noche se acostaba a medianoche.

The words underlined in the English version (a) correspond to the underlined forms in the Spanish version (b). The mismatch between these two verb phrases is readily apparent. The English uses three times as many words as the Spanish, including one repeated item TO. The English uses a complex auxiliary construction USED TO and a periphrastic verb form GO TO BED, where Spanish uses only a single lexical item and a reflexive pronoun. Clearly, deleting ACOSTABA from the Spanish sentence removes from the passage a large number of semantic and syntactic features which all have to be supplied by the reader. None of the English words would place such a large load on the reader if they were deleted.

To some extent, this heavy load borne by Verbs in Spanish is compensated for by the rules of Noun Phrase Concord. Thus, in examples (c) and (d), possible answers for the gap in the Spanish sentence are restricted by both the preceding definite article and the following adjective, both of which are marked as feminine:

- c) una \_\_\_\_\_ construída a la orilla de un río
- d) a \_\_\_\_\_ built on the bank of a river

Any of the following words would fit the English: house, castle, town, building, city, church, wall, besides many others, but of their Spanish translations only the uppercase forms could fill the gap because of the gender restrictions CASA, castillo, pueblo, edificio, CIUDAD, IGLESIA, MURALLA. In theory, this should make it easier for the reader to find the word intended by the author, since the choice of acceptable words is limited and cut by about half. In practice, however, it seems probable that constraints of this kind actually make it harder for a learner to find any appropriate item in the limited vocabulary available to him. In the example above,

for instance, it is fairly clear that some form of building is intended, and in the English test, as long as this was realized, the superordinate term BUILDING could be used appropriately. In Spanish, however, the superordinate term EDIFICIO is masculine and therefore not acceptable on morphological criteria, even though it is perfectly satisfactory on semantic grounds. A subject who realised that a building was required must therefore choose a particular instance of a building, and run a high chance of making an inappropriate choice.

All these factors combine together to make a Cloze Test in Spanish rather more difficult than an "equivalent" passage in English and under these circumstances it is not clear that we can safely assume that the correlations between "overall competence" and Cloze Test scores found by Oller for English could be expected to be found also in Spanish.

(For a further discussion of some of these ideas see Freeland and Meara (in prep.)).

The second major problem with Cloze Tests in languages other than English is the question of cultural norms. Cloze Tests in English seem to produce a relatively narrow range of responses to any given gap, and native speakers would normally be expected to achieve a score of 60-80% on a non-technical passage scored according to the exact word criterion. Both these characteristics are important contributors to the reliability of Cloze Tests as a measure of overall ability in a foreign language. Suppose, however, that one was dealing with a language in which non-technical texts produced a much wider range of responses in native speakers. This would reduce the expected native speaker score to a much lower level, say in the 30-50% range, and it is unlikely that scores as low as this would still discriminate reliably between native speakers and good learners.

This question of cultural norms is one that has received little attention so far. However, it seems plausible to suggest that Spanish passages could well be expected to elicit a rather wider range of responses than comparable English passages. English vocabulary falls into two rather clearly defined classes: words that are etymologically related to Latin and Greek roots, and those which are etymologically related to Germanic roots. The former tend to be restricted in use, occurring predominantly in certain registers, which are thus heavily marked in style. In Spanish this dichotomy is much less clear, and there is much greater flexibility in the use of individual words. (This notion may account to some extent for Rosenzweig's finding that word association responses in other Romance Languages (French and Italian) are much less stereotyped than responses produced by English and American speakers (Rosenzweig (1969)) and the indications are that similar stereotypy characterizes Spanish (Haworth (1979)). If low scores such as these should turn out to be typical of Cloze Tests in Spanish, then the reliability of the tests as a measure of overall competence in Spanish would be seriously reduced.

In short, then, there are some serious gaps in our knowledge of the way Cloze Tests in Spanish might work, and without this necessary background information it seems premature to use Cloze Tests as a way of providing anything more than an indication of a minimal level of competence in Spanish.

These methodological considerations aside, there is one other reason for not using independent tests of the learners' overall ability. Apart from using such tests to discriminate between the different learner groups, there is little else that could be done with data of this sort. It might be argued that it would be

possible to correlate these scores with the individual scores obtained in the experimental tasks, with the expectation that learners who score highly on the overall ability tests will perform more like native speakers on the experimental tasks. This argument seems plausible, but it runs into the difficulty that it is not always clear what scores one would want to correlate in this way, or what the results of such a correlation might mean. For example, in the case of our Tachistoscopic Recognition Thresholds experiment, it is by no means obvious what scores one could use in this way. Furthermore, the argument assumes that there is some kind of regular development in the way that second language speakers come to make use of syntactic structures, and that this developmental pattern will interact with overall language ability in some easily definable way. Though this seems to be a plausible assumption, there is actually no evidence to support it. Indeed, as we have seen from our discussion of the work of Bruce and Pugh (1966) with children there is some evidence that using syntactic structure is not a characteristic that grows steadily over time. Rather there is a sudden rapid change in behaviour patterns around the age of seven years. Prior to this change, children show no evidence of structuring behaviour; after the change they perform like adults; the change itself, however, appears to occur within the space of a few months. If a similar catastrophic change could be located in second language learners, this would be of considerable interest, but there is no obvious way of relating children's ages and learners' Cloze Test scores, and this makes it difficult to guess where one might start looking for such a change in learners. Cloze Test scores provide only an



ordering of the subjects relative to each other, and not an objectively defined account of the state of their ability in the second language.

For these reasons then, the classification of the learners used in the experiments was felt to be adequate, if not wholly satisfactory. It is clear, however, that there is a strong case for developing a research tool based on Cloze Tests which could be used in conjunction with the type of research that has been advocated here. Such a tool would allow us to ask questions that are much more specific than those asked in this thesis. In particular, a useful development would seem to be a set of about twenty Cloze Tests in Spanish (or any other language) for which the responses of a large number of native speakers are available. Such data would allow us to resolve our doubts about problems such as the expected range of native speakers scores. More importantly, they would allow us to give subjects a score based not on the performance of other learners, but on the expected average score of the native speaker group, and this would allow us to refine our classification of learners to a degree which is just not possible with the experimental tools that are currently available.

Table 7-B-1

Lea el párrafo que sigue y en cada espacio subrayado escriba una palabra que le suena correcta.

Por la mañana Pedro había experimentado una de las sensaciones más grandes de su vida: ver el mar por primera vez. Al salir él de la casa llovía, \_\_\_\_\_ (1) eran tan finas las gotas y \_\_\_\_\_ (2) espaciadas que las personas que llevaban \_\_\_\_\_ (3) lo llevaban cerrado. ¿Qué lluvia más \_\_\_\_\_ (4)! En su tierra o la lluvia \_\_\_\_\_ (5) las calles, o granizaba o nevaba \_\_\_\_\_ (6) hacía sol; pero esta clase de \_\_\_\_\_ (7) no lo había visto nunca. Caminó \_\_\_\_\_ (8) la plaza y torció a mano \_\_\_\_\_ (9), por la primera calle. Al fondo \_\_\_\_\_ (10) unos jardines con unos pinos muy \_\_\_\_\_ (11), cuyas ramas formaban como un techo, \_\_\_\_\_ (12) el que paseaba la gente. Le \_\_\_\_\_ (13) tanto la forma y el color \_\_\_\_\_ (14) los árboles que tardó en ver \_\_\_\_\_ (15) breve línea azul que se veía \_\_\_\_\_ (16) los trancos. Aceleró la marcha y \_\_\_\_\_ (17) acercó, y ya no vió los \_\_\_\_\_ (18) ni pensó en la extraña lluvia, \_\_\_\_\_ (19).  
acababa de descubrir el mar.

## ☛ Second Language Learners and Other Non-Standard Populations.

Throughout this discussion we have been at some pains to stress that the results of the non-native speaker groups studied were contradictory and unsystematic. In some of the experiments, results were obtained which clearly showed that appropriate structural effects were produced by the stimulus material used, and that these effects were smaller in the case of the learners than in the case of the native speaker subjects. Chapter 2 and Chapter 5 illustrate this - indeed it will be recalled that in Chapter 5 there was a total absence of a structure effect in the non-native subjects. Chapter 3 and Chapter 6, however, showed that what had originally been believed to be a simple effect ascribable to the influence of syntactic structure, may actually have been more complex than it appeared, since in these experiments, contrary to prediction, the results of the learner groups were apparently more structure-sensitive than the native speakers. The simplest explanation of these latter findings is probably that the experimental tasks are poorly understood and that the standard explanation of the structural effects found with native speakers is an incorrect one.

If we are prepared to accept that some other explanation may account for the odd results in these experiments, then it is possible that the findings may be rather less anomalous than they appear at first sight. For some of the experiments, data obtained from comparable studies using subjects who are not normal adult native speakers is available, and this data suggests that there is some similarity between the performance of such groups and the performance of the

learner groups studied here. Directly comparable data of this type is available for two of the experimental tasks - recall of statistical approximations and the probe latency task - and some other evidence is also available, which though not directly comparable with our tachistoscopic recognition experiment, does allow us to make an educated guess at how subjects who are not normal adult native speakers might be expected to behave on this task.

#### A. Statistical Approximations.

Experiments using Statistical Approximations have been carried out on two main non-standard populations: Children and Schizophrenics.

The chief source of data for children is a paper by Bruce and Pugh (1966) which tested the ability of six year old children to handle strings of words at various orders of approximation in an experimental task closely comparable to the one reported in Chapter 2. Bruce and Pugh constructed a set of approximations to English, using Miller and Selfridge's method, but using the children themselves to generate their own strings as a way of controlling for syntactic and lexical complexity and thus ensuring that the material was not in principle beyond the ability of the children. The stimulus strings are thus not really approximations to English, but rather approximations to Six-Year-Old English. These strings - of seven and ten words - were presented aurally to the children, who were asked to recall them. Bruce and Pugh's results are complex. The group recall-patterns for the various orders of approximation show some evidence of a structural effect, with higher orders of approximation being recalled slightly better than lower orders, but, given that the strings used were very short, the effect is small, and far below

the level of performance that would be expected of an adult. Bruce and Pugh showed, however, that this analysis was a superficial one, in that it masked a clear dichotomy in the subjects. Some of these individually performed in a manner that was clearly like the native speaker pattern of responses, while others showed essentially flat response curves, which did not suggest any sensitivity to syntactic structure. Bruce and Pugh went on to show that the more linguistically advanced children tended to perform in closer accordance with the adult model. They further showed in a subsequent retest that a large proportion of the children with flat curves had apparently changed their pattern of responding and begun to produce curves more closely resembled the adult models. This finding was interpreted as evidence for a sudden catastrophic change in the children's sensitivity to syntactic structure at about the age of six years.

Figure 7C-1 below shows examples from Bruce and Pugh of a typical adult like curve from one child, and a typical non-adult set of scores from a different child.

Figure 7C-2 shows the mean group scores for the first and second tests, and for the purposes of comparison the mean score of the less advanced group of learners studied in our experiment reported in Chapter 2. This comparison is not totally problem free, of course, since the two experiments used different materials, Bruce and Pugh English and ours French, but the two sets of data are not wholly dissimilar. The learner scores are in the same general area as those of the native speaking children. Furthermore, some of the subjects who were singled out for special mention in Chapter 2, showed response curves that are very similar to the patterns produced by children with lower levels of linguistic development.

Figure 7C-1.

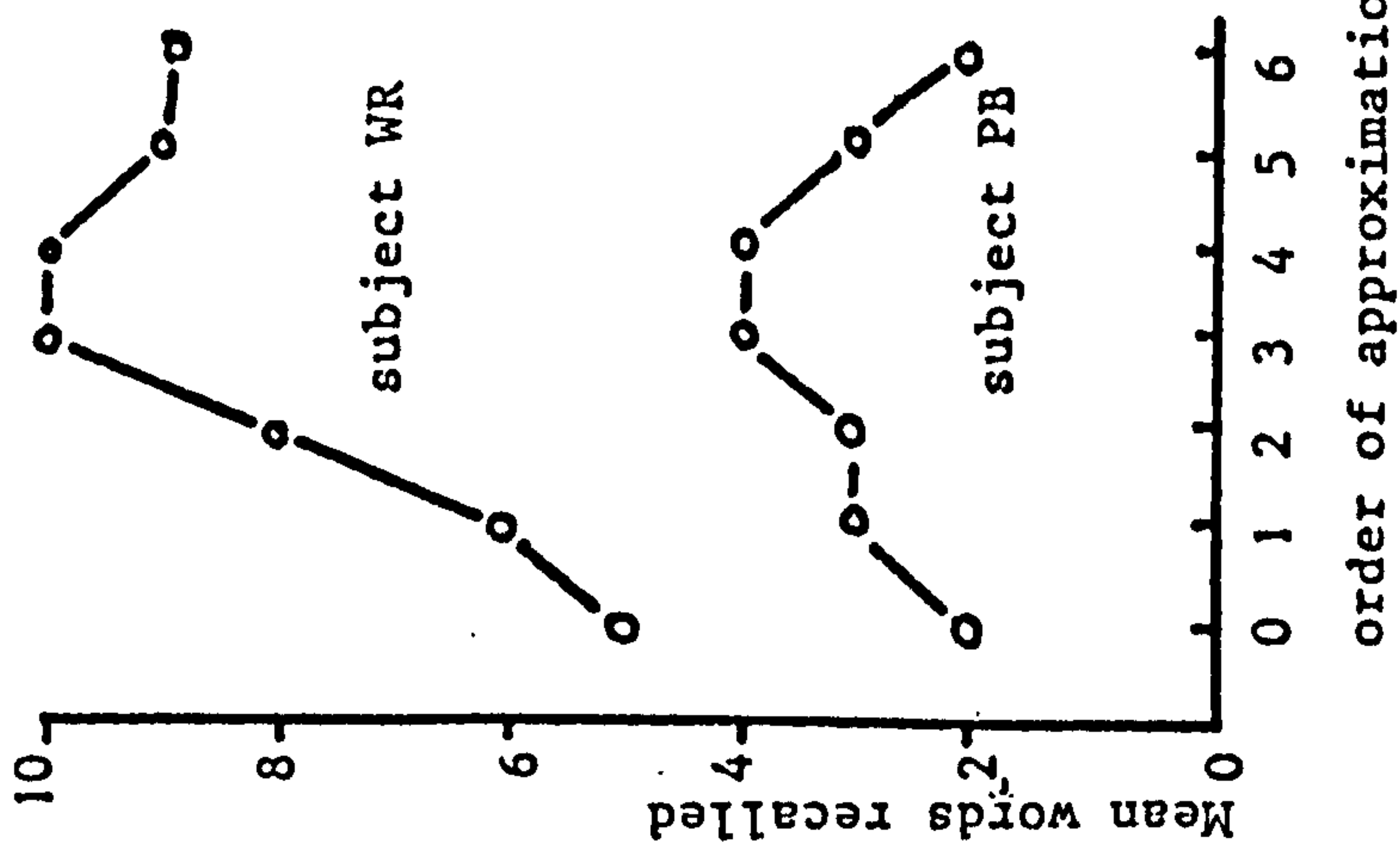


Figure 7C-2.

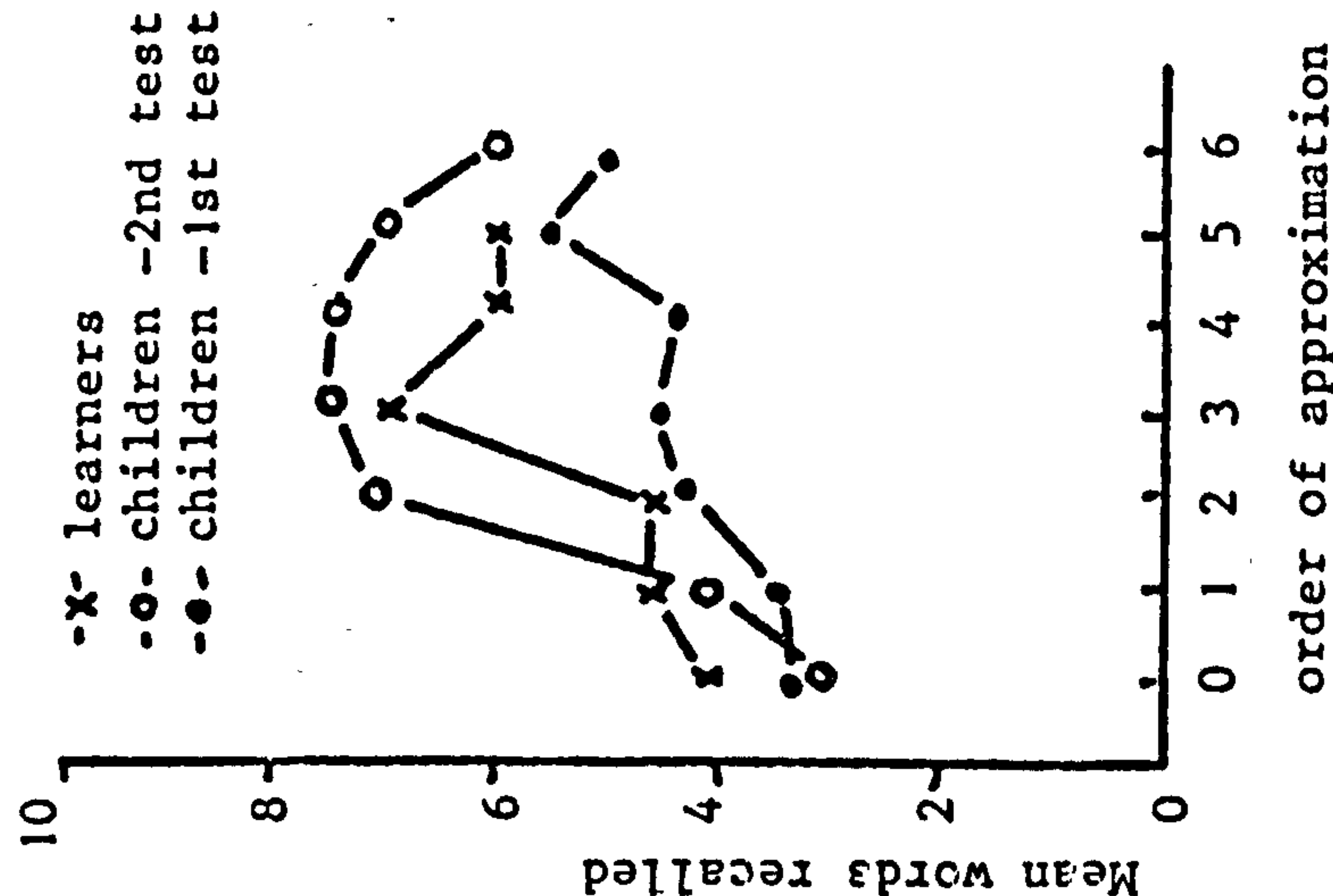
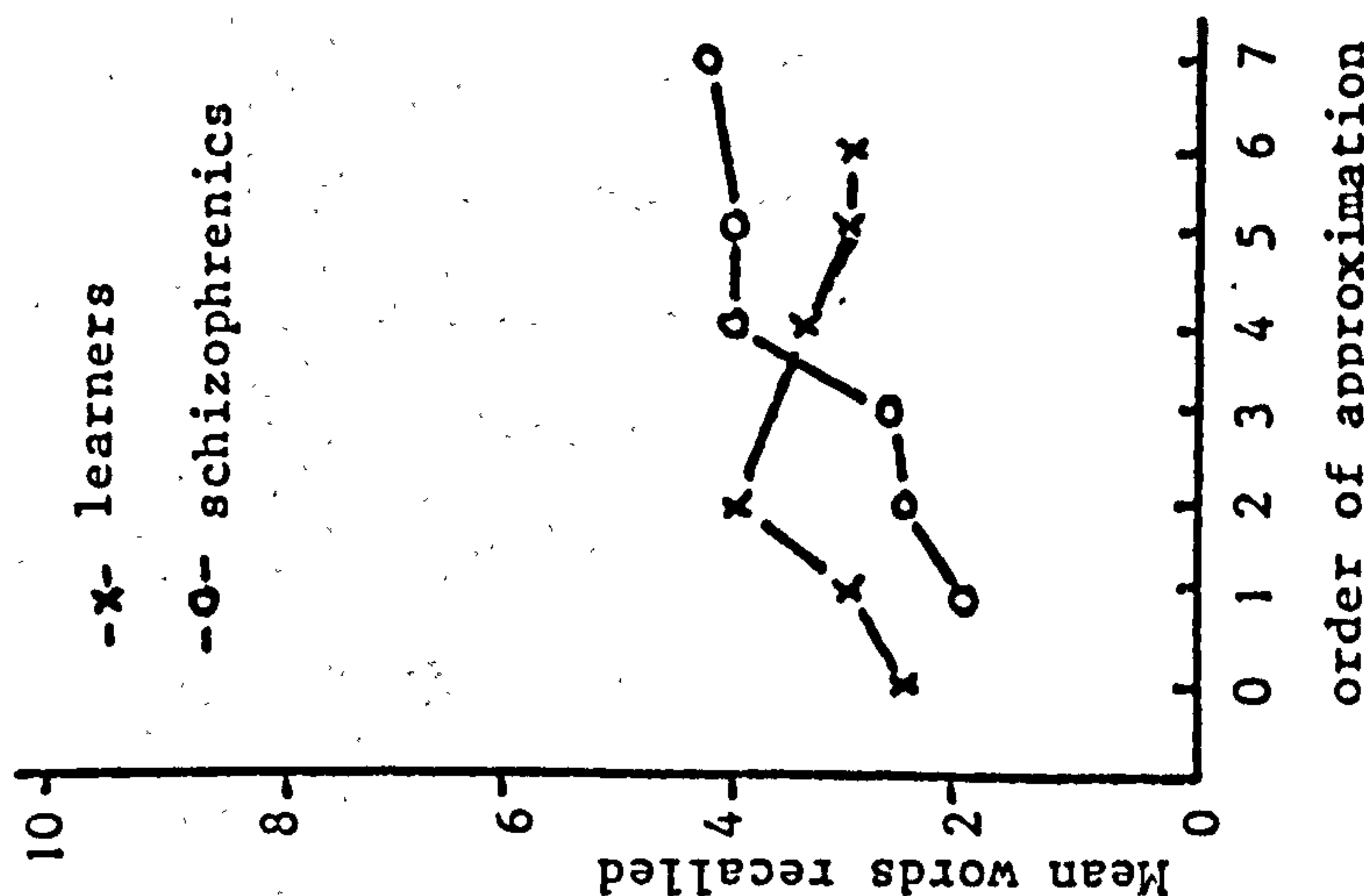


Figure 7C-3.



A comparable degree of similarity can be found between the learner groups of Chapter 2 and the results of schizophrenic patients on tasks involving statistical approximations. Schizophrenics are widely believed to be insensitive to syntax - though in the light of the problems we have encountered in attempting to test this claim for second language speakers, this belief should perhaps be regarded as an oversimplification and treated with appropriate caution. Statistical approximations were widely used in the sixties as a way of testing the claim, however, and generally the results of these tests show quite marked differences from the results of tests using normal adult native speakers as subjects, and a close similarity to those of the children studied by Bruce and Pugh.

Figure 7-3 shows a set of results of this sort from Lawson, McGhie and Chapman (1964) which illustrate this point. Again, for the purposes of comparison, data from our Chapter 2 is repeated to show the extent of the similarity between the learners' scores and those produced by the schizophrenics. Again, the comparison between the two sets of data is not wholly satisfactory, since different materials were used, but making allowances for this, there is some similarity between the two sets of scores, which differ from the standard normal native speaker scores in much the same way.

#### B. Probe Latencies.

There is no data on probe latencies for schizophrenic subjects - though a number of other similarities between second language speakers and schizophrenics suggest that the results discussed above may not be an isolated phenomenon. (cf. Meara (1978) for a further detailed discussion of this idea.) However, data on probe latencies for

children are available, and one such set of results was discussed in Section 6-1. Here, it will be recalled that the children's responses were generally rather slower than the adult responses, and showed an even more marked syntactic effect than the adults did, with the longest latencies clearly occurring when the response and the probe words straddled a major constituent boundary.

This result ought to be something of an embarrassment to Suci Ammon and Gamlin, who first reported the findings, as the findings seem to imply that children are actually more sensitive to syntactic structure than adults are, where one might reasonably have expected the contrary result. Suci, Ammon and Gamlin do not discuss this inconsistency, however. Possibly we have here a task whose explanation is less well understood than is at first apparent from Suci, Ammon and Gamlin's account, and one where comparisons between adults and children may be theoretically simple and straightforward, but much less so in practice. In section 7-2-3, we argued that the "same" task given to two different groups of subjects, might not in fact be the same task at all if basic controls failed to eliminate unwanted variables that adversely affected one group of subjects but not the other. Whatever the explanation of this result, however, the strikingly large differences in responses at different probe positions, and the clear relationship between these differences and the syntactic structure of the stimulus sentences in both Suci, Ammon and Gamlin's children and the learners studied here should be obvious.

### C. Recognition Thresholds and Click Placement.

Experimental studies of click placement and recognition thresholds for word pairs presented tachistoscopically do not appear to



have been carried out with non-adult or non-normal populations. It seems reasonable to predict, however, that if such studies were carried out with children, for example, their results would not be altogether different from those found here with second language learners. Consider first recognition thresholds. Children clearly do go through a stage before they become fluent readers when they read text word by word rather than in the larger phrase-sized chunks used by skilled, experienced readers, and they experience some difficulty in recognizing words presented tachistoscopically (Vernon (1971), esp. pp28-50). Such children would presumably fail to make use of syntactic information available in brief visual displays, and would thus fail to discriminate between structured and unstructured word sequences just as the learner groups did. The fact that children's eye-voice span is much smaller than that of skilled adult readers lends some further support to this speculation (Buswell (1970)) (Interestingly, recent work by Bell (1979) has suggested that second language speakers also have eye-voice spans that are much smaller than native speakers eye-voice spans even when the learners are relatively advanced in their second language.) Similar considerations hold in the case of click location. In Chapter 3 it was argued that the mislocations of clicks could be partly explained in terms of difficulties in word recognition, displacements generally being leftwards if word recognition took more than a very small length of time, and the larger the longer recognition was delayed after presentation of the word in which the click occurred. It seems plausible to suppose that children recognize words more slowly than adults, and it follows that they should therefore produce larger mislocations of clicks than adults do.<sup>19</sup>

The same arguments might also be advanced in the case of schizophrenic subjects. These subjects are known to be much worse at recognising auditorily presented words than native speakers are (Bull and Venables (1974)) which would lead one to predict that schizophrenics too would have wider mislocations of clicks than normal subjects. Furthermore, schizophrenics are less effective in their handling of visual arrays of verbal material than normal subjects, and this difficulty increases with a larger display (Neale and Cromwell (1977)). These latter findings suggest that schizophrenic subjects would find considerable difficulty with the tachistoscopic recognition task used here, and that they would be unlikely to show lower thresholds on syntactically structured material.

D.

It appears, then, that the inconsistencies in the learner results reported in the previous chapters may not be as unsystematic as they appear to be. Results from comparable studies with other non-standard populations are in line with our findings, and where such data are not available a plausible case can be made out for the claim that quite disparate non-standard populations might be expected to produce results not entirely dissimilar from those found with second language speakers.

While it would be foolish to argue that these similarities are all ascribable to identical causes, they do suggest that the problems encountered by second language learners may be of wider general interest than is usually supposed. The most likely explanation of the similarities in the behaviour of these disparate groups is that there are some quite serious constraints on the way normal language

behaviour is likely to break down, and that abnormal language behaviour, whether it is the result of pathological causes of the schizophrenic type, or whether it arises because of incomplete learning as is the case with the children or the second language learners, differs from the standard of normal language behaviour in only a limited number of ways. If this were true, then studying the psychology of speaking a second language might turn out to be a valuable source of insights into both the problems of language development in children, and the more complex problems posed by pathological conditions of language.

FOOTNOTES

Footnote 1.

page 7. Derwing criticises Chomsky's distinction between Competence and Performance on the grounds that it is obscure and confusing. He distinguished three mutually inconsistent interpretations of Chomsky's position:

a) Competence as an idealized model of linguistic performance. Derwing criticises this interpretation on the grounds that the experimental evidence which claims to support this view is all unsound. He further argues that because Chomskyian Grammars are essentially random generators of sentences, they are intrinsically incapable of serving as models of performance because they cannot handle the important pragmatic considerations that affect ordinary language behaviour.

b) Competence as a Central Component of an idealized Performance model. Here Derwing argues that no explicit and testable performance model which incorporates a competence model as a proper sub-component exists, and he disputes the logic that argues for the necessity of building competence models before the construction of performance models can be attempted.

c) Competence as an independent abstract entity remote from linguistic performance. Derwing feels that this is the only interpretation of Competence which stands up to scrutiny, but he argues that the implications of such a position for the study of performance are by no means as clear as is generally supposed.

Derwing's work is considered by Smith (1975) who represents the main-stream transformationalist viewpoint, as "tendentious and often incoherent".

Smith does acknowledge that Derwing's discussion of Competence and Performance in Chomsky's work is accurate.

For further discussion of the Competence/Performance question see Clark (1974).

Footnote 2

page 16 He might have used a measure of concordance (e.g. Kendall's W) or either of the two more recently developed methods of scoring data of this type devised by Dulay and Burt (1975) (though cf Rosansky (1976) for a full discussion of the problems of scoring developmental sequences).

Footnote 3

page 37 The main difference between these two claims is one of character. The Reduced Redundancy Conjecture is really a description of learner behaviour, and not an explanation of it. It claims only that learners behave as if the redundancy of Foreign Language material were reduced, and offers no explanation for this effect. The Word List Conjecture, on the other hand, makes a stronger explanatory claim, and is thus to be preferred because it is more clearly refutable than the Reduced Redundancy Conjecture.

Footnote 4

page 54 Jumbled sentences are unsatisfactory because random jumbling often leaves constituents intact or creates new constituents by chance. Furthermore, in a gender marked language such as French or Spanish, random jumbling often juxtaposes elements which are incorrectly matched for gender. This latter effect causes severe interference problems for native speakers. Lexically dense stimulus strings are discussed in Chapter 4.

Footnote 5

page 75 The sex of the learners may be relevant here because of the well-known propensity of girls to be better language learners than boys of the same age.

Footnote 6

page 96 I take this to mean that Coleman finds the asymptote in Miller and Selfridge's results is problematical. If scores fail to rise beyond, say, sixth order of approximation, this seems to imply that syntactic relationships which extend over sequences of eight, nine, ten or more words are essentially irrelevant as far as short term memory is concerned, and not implicated in the mechanisms that underlie the storage processes. Coleman obviously feels that this inference is false.

Footnote 7

page 110 In order to simplify the discussion, I am adopting the standard conventions used to describe mislocations, Thus, clicks which are recorded as occurring before their objective position are referred to as having moved left-wards. Clicks which are perceived to have occurred after their objective occurrence are said to have moved right-wards. This spatial metaphor is used as a convenient short-hand, and it is not intended to lose sight of the fact that the displacement is actually of a temporal sort.

Footnote 8

page 112 It is unsatisfactory because it predicts that displacement should decrease towards the end of the list of digits, since information is steadily reduced by the presentation of each succeeding digit. There is no support for this prediction in the data, however. Furthermore, the claim predicts that displacement in lists of digits should be smaller than displacement in lists of words. There is no support for this claim either.

Footnote 9

page 121 This is an example of a problem to which we shall return in Section 7-3. Clearly Maidment is assuming that the click paradigm can be transferred in toto to experiments involving second language learners. In fact, however, the technique clearly depends on the ability of the subjects to understand what they hear. This ability can be largely taken for granted with native

speakers, of course, but with learners the assumption is clearly much less plausible. In this case, Maidment's materials make it very unlikely that her least advanced learners were able to understand the stimuli at all. If the learners were not able to do this (and Maidment states explicitly that they were not expected to understand the stimulus sentences) then their task is essentially different from the task facing a native speaker, and all basis for comparison between learners and native speakers ceases to exist.

Footnote 10

page 165 The argument here seems to be that if Roman script with its left-right orientation goes hand-in-hand with a right visual field effect (i.e. a left hemisphere superiority), then Semitic scripts with their right left orientation might produce a contrary left visual field effect. (i.e. a superiority for material presented to the right hemisphere. One could argue, however, that a right-left orientation ought to produce an enhanced right visual field effect instead.

Footnote 11

page 167 The point here is that the correlation figure tells you about individuals, but without more details it is not possible to infer from this that there is any kind of systematic difference between scores in an  $L_1$  and an  $L_2$ .

Footnote 12

page 175 R. Clark has pointed out to me that an alternative explanation may lie in the scores of the individual subjects, if some of them show an improvement while others do not. This is clearly a strong reason for looking at individual scores, instead of treating learners as part

as part of a mass group of subjects. This point also bears on the discussion in Section 7-3, where it could be argued that assumptions made about homogeneity in native speakers may be wholly unapplicable to non-native speaker groups, and that very large groups of non-native speaker subjects may be required for reliable experimental data to be found.

Footnote 13

page 181 This is not meant to imply that low-level syntactic structures are unimportant. Quite the contrary, in fact, since the Revised Word List Conjecture ascribes greater importance to these structures than is customary, and implies that learners rely on these structures much more heavily than native speakers do.

Footnote 14

page 183 This chapter is concerned primarily with recognizing printed words, and therefore this discussion concentrates on technical problems that arise when printed words are used as stimuli. By extension, however, it also raises problems of comparability in experiments using two languages and an auditory mode of presentation. (cf. Section 7-4 for a further discussion of this problem.)

Footnote 15

page 196 This variable was not controlled for in the original choice of subjects, since there is a general assumption that, for all practical purposes, there is no difference between adults in the range of 15-40, say. However the results of a subsequent experiment, reported in Section 5-4-5, where older subjects were used show results very similar to those of the young



adults studied here. This strongly suggests that the explanation on the basis of age is false, and that the group effects found here are something to do with the native language of the subjects. This notion supports the suggestions put forward on page 197.

Footnote 16

page 222 K and W comment "when the final word of a sentence occurred as a signal subjects were instructed to say as quickly as possible 'nothing'." This procedure seems so absurd that comment is superfluous.

Footnote 17

page 243 For each subject, the median of the latencies for his correct responses was used as the score in these analyses. The graphs show the means of these median scores at each probe position. S. A and G do not comment on why median scores rather than mean scores were used, but since the chief difference between them is that the median is less effected by extreme scores or wide variation, it seems reasonable to assume that this figure was adopted in order to reduce the overall variation of the scores. It is actually common practice among people using latency techniques to subject the data to an arc sin transformation before carrying out the final analysis. This transformation also has the effect of reducing inhomogeneity of variance. The new experimental work reported in this chapter follows the procedure of using medians as raw data, but does not use transformations of any kind. In fact, however, using the means instead of the medians does not produce any major change in the pattern of the results.

Footnote 18.

page 293

This is not meant to imply that there are no differences between language development in different language communities. I merely wish to argue that sufficient work has been done in this field for the claim that there are universals in development to be one which needs to be taken seriously.

Footnote 19

page 353

Ruth Clark has pointed out to me that some of these claims are tested in a paper by Seitz and Springer-Bloom (1977). This paper studied the effects of age on click migration effects by asking children of different ages and adults to carry out a standard click location task. The results of this study showed

a) that click placement was more accurate in older subjects than in younger ones, in that a higher proportion of clicks were correctly assigned by the adults than by the children; and

b) that migrating clicks were differently affected by syntactic structure in children and adults, in that clicks not located at a major constituent boundary were more likely to migrate towards it in the adults' judgements than in those of the children.

The first of these findings is clearly in line with the argument developed in this section. The second finding is more problematical, in that the figures are reached by an averaging procedure comparable to that used by Ladefoged and Broadbent and criticised earlier in section 3-2. In the case of the youngest children the standard deviations of the scores are very large, and this suggests that the use of mean scores may be particularly misleading here. There is also a suggestion that the younger children's scores are misleading in that none of them; apparently, indicated that clicks might have occurred between words. This happens very frequently with adults.

Bibliography:

Albert M. and L. K. Obler. 1979. The Bilingual Brain. Academic Press.

Ammon P. 1968. "The Perception of grammatical relations to sentences a methodological exploration." Journal of Verbal Learning and Verbal Behaviour. 7.869-875.

Anderson J. 1976. Psycholinguistic Experiments in Foreign Language Testing. University of Queensland Press.

Aronson Berman R. 1978. "Contrastive Analysis Revisited." Inter-language Studies Bulletin. Utrecht. 3.2.212-234.

Awbery G. 1976. The Syntax of Welsh. Cambridge University Press.

Bach E. 1974. Syntactic Theory. Holt Rinehart and Winston.

Bach E. and R. T. Harms. 1968. Universals in Linguistic Theory. Holt Rinehart.

Barton M, H. Goodglass and A. Shai. 1965. "Differential recognition of tachistoscopically presented English and Hebrew words in right and left visual fields." Perceptual and Motor Skills. 21.431-437.

Bayes. H. 1969. unpublished work. Department of Linguistics, University of Cambridge.

Bell, T. unpublished work. 1979.

Bennett W. A. 1974. Applied Linguistics and Language Learning. Hutchinson.

Bever T. and D. T. Langendoen. 1971. "A Dynamic Model of the evolution of language." Linguistic Inquiry. 2.433-463.

Burt M.K., H.C. Dulay and E. Hernandez Ch. 1975. Bilingual  
Syntax Measure. Technical Handbook. Harcourt Brace Jovanovich.

- Bousfield W. A. 1953. "The occurrence of clustering in the recall of randomly arranged associates." Journal of General Psychology. 49.229-240.
- Brener R. 1940. "An experimental investigation of memory span." Journal of Experimental Psychology. 26.467-482.
- Brown R. 1974. A First Language: the Early Stages. Allen and Unwin.
- Brown R. and D. McNeill. "The tip of the Tongue Phenomenon." Journal of Verbal Learning and Verbal Behaviour. 5.325-337.
- Bruce D. J. and H. M. Pugh. 1966. "Immediate Verbal Memory and Linguistic sophistication in a group of six year old children." Language and Speech. 9.2.69-83.
- Bruner J. S. and D. O'Dowd. 1958. "A note on the informativeness of parts of words." Language and Speech. 1.98-101.
- Bull H. C. and P. H. Venables. 1974. "Speech Perception in Schizophrenia." British Journal of Psychiatry. 125.350-360
- Buswell G. T. 1920. An experimental study of eye-voice span in reading. Supplement to Educational Monographs 17. University of Chicago Department of Education.
- Carroll J. B. and S. M. Sapon. 1959. Modern Language Aptitude Test. New York.
- Carrow E. and M. Mauldin. 1973. "Children's recall of approximations to English." Journal of Speech and Hearing Research. 16.201-212.
- Cattell. J. 1885. "Experiments on the Association of ideas." Mind. 12.68-74.
- Chomsky N. 1965. Aspects of the Theory of Syntax. MIT Press.
- Chomsky N. and G. A. Miller. 1963. "Introduction to the formal analysis of natural languages." in Luce et al.

Clark R. 1974 "Performance without Competence." Journal of  
Child Language. 1.1-10

Cook V.J. 1979. "Aspects of Memory in Secondary School Language  
Learners." Interlanguage Studies Bulletin. 4.2.161-172

- Chomsky N. 1965. Aspects of the Theory of Syntax. MIT Press.
- Chomsky N. and G. A. Miller. 1963. "Introduction to the formal analysis of natural languages." in Luce et al.
- Clark H. H. and E. V. Clark. 1977. Psychology and Language. Harcourt Brace Jovanovich.
- Clarke M. A. 1976. "Second language acquisition as a clash of consciousness." Language Learning. 26.2.377-390.
- Cofer C. N. 1961. Verbal Learning and Verbal Behaviour. John Wiley
- Cohen A. 1975. "Forgetting a second language." Language Learning. 25.1.127-138.
- Coleman E. B. 1963. "Statistical Approximations to English: some comments on the method." American Journal of Psychology. 76.239-247.
- Coltheart M. 1977. Review of Gibson and Levin: The Psychology of Reading. Quarterly Journal of Experimental Psychology. 29.157-167.
- Cook V. J. 1978. "Second Language Learning: a psycholinguistic perspective." Language Teaching and Linguistics Abstracts. 11.2.73-89.
- Corder S. P. 1967. "The significance of learners' errors." International Review of Applied Linguistics. 5.161-170.
- Corder S. P. 1971. "Idiosyncratic dialects and error analysis." International Review of Applied Linguistics. 9.2.147-159.
- Corder S. P. 1973. Introducing Applied Linguistics. Penguin.

- Clark H. H. and E. V. Clark. 1977. Psychology and Language.  
Harcourt Brace Jovanovich.
- Clarke M. A. 1976. "Second Language Acquisition as a clash of  
consciousness." Language Learning. 26.2.377-390.
- Cofer C. N. 1961. Verbal Learning and Verbal Behaviour. John Wiley.
- Cohen A. S. 1975. "Forgetting a Second Language." Language Learning.  
25.1.127-138.
- Coleman E. B. 1963. "Statistical Approximations to English: some  
comments on the method." American Journal of Psychology.  
76.239-247.
- Coltheart M. 1977. Review of Gibson and Levin: The Psychology of  
Reading. Quarterly Journal of Experimental Psychology.  
29.157-167.
- Cook V. J. 1978. "Second Language Learning: a psycholinguistic  
perspective." Language Teaching and Linguistics Abstracts.  
11.2.73-89.
- Corder S. P. 1967. "The significance of Learners' Errors."  
International Review of Applied Linguistics. 5.161-170.
- Corder S. P. 1971. "Idiosyncratic dialects and error analysis."  
International Review of Applied Linguistics. 9.12.147-159.
- Corder S. P. 1973. Introducing Applied Linguistics. Penguin.
- Corder S. P. 1977. "Error analysis, Interlanguage and Second  
Language Acquisition." in Kinsella V. 1978.
- Craik F. I. M. and P. A. Masani. 1967. "Age differences in the  
temporal integration of language." British Journal of  
Psychology. 58.291-299.



- Dalrymple-Alford E. C. 1967. "Prestimulus language cuing and speed of identifying Arabic and English words." Psychological Reports. 21.27-28.
- Dato D. P. 1969. "The development of the Spanish Verb Phrase in children's second language learning." in Pimsleur and Quinn. 1971.
- Dato D. P. 1970. American children's acquisition of Spanish syntax in the Madrid environment. United States Office of Health, Education and Welfare.
- Dechert H. 1979. Pausological Implications of Speech Production. Mouton.
- Deese J. 1961. "From the isolated Verbal Unit to connected discourse." in Cofer. 1961. pp11-31.
- Deese J. and R. A. Kaufman. 1957. "Serial effects in recall of unorganized and sequentially organized verbal material." Journal of Experimental Psychology. 54.180-187.
- Derwing, B. L. Transformational Grammar as a theory of Language Acquisition. Cambridge University Press.
- Dixon T. R. and Horton D. L. 1968. Verbal Behaviour and General Behavioural Theory. John Wiley.
- Dornic S. 1969. "Verbal factor in number perception." Acta Psychologica. 29.393-399.
- Dulay H. and M. Burt. 1974. "Natural Sequences in Child Second Language Acquisition." Language Learning. 24.1.245-253.
- Ellis A. 1979. "Speech Production and Short Term Memory." in Morton J. and J. Marshall. 1979

- Epstein W. 1961. "The influence of syntactical structure on learning." American Journal of Psychology. 74.80-85.
- Epstein W. 1962. "A further study of the influence of syntactical structure on learning." American Journal of Psychology. 75.121-126.
- Ervin S. M. 1961. "Changes with age in the verbal determinants of word association." American Journal of Psychology. 74.361-372.
- Fay D. and A. Cutler. 1977 "Malapropisms and the structure of the mental lexicon." Linguistic Inquiry. 8.505-520.
- Ferguson C. and D. Slobin. 1973. Studies of Child Language Development. Holt Rinehart and Winston.
- Fillmore C. J. 1968. "The Case for Case." in Bach and Harms. Eds. 1968.
- Flesch R. F. 1948. "A new readability yardstick." Journal of Applied Psychology. 37.221-223.
- Flores d'Arcais G. B. and W. J. M. Levelt. 1970. Advances in Psycholinguistics. North-Holland.
- Fodor J. A., T. G. Bever and M. F. Garrett. 1974. The Psychology of Language. McGraw-Hill.
- Fodor J. and M. Garrett. 1967. "Some syntactic determinants of sentential complexity." Perception and Psychophysics. 2.289-296.
- Forster, K. I. 1970. "Visual Perception of rapidly presented word sequences of varying complexity." Perception and Psychophysics. 8.215-221.
- Foulke E. 1969. "The perception of time-compresses speech." in Horton and Jenkins (eds.) 1969.
- Freeland J. and P. Meara (in prep.) A Cloze Test Reader.

- Garrett M., Bever, T. A. and J. Fodor. 1966. "The active use of grammar in speech perception." Perception and Psychophysics. 1.30-32.
- Garvey W. D. 1953. "The intelligibility of speeded speech." Journal of Experimental Psychology. 45.102-108.
- Gerver D. 1969. "Effects of grammaticalness, presentation rates and message length in auditory short term memory." Quarterly Journal of Experimental Psychology. 21.203-208
- Gibson E. J. and L. Guinet. 1971. "The perception of inflections in brief visual presentations of words." Journal of Verbal Learning and Verbal Behaviour. 10.182-189
- Gibson E. J. and H. Levin. 1975. The Psychology of Reading. MIT Press.
- Glanzer M. 1962. "Grammatical category: a rote learning and word association analysis." Journal of Verbal Learning and Verbal Behaviour. 1.34-41.
- Gougenheim G. 1958. Dictionnaire Fondamental de la Langue Française. Harrap-Didier. Paris.
- Green J. 1976. "How free is word order in Spanish?" in Harris M. ed. 1976.
- Greenberg J. 1966. Language Universals. Janus Linguarum Series Minor. 59. Mouton.
- Greene J. 1972. Psycholinguistics: Chomsky and Psychology. Penguin
- Grosjean F. 1978. "Linguistic Structures and Performance Structures: studies in Pause Distribution. in Dechert (ed.) 1979.
- Hakes D. T. and D. J. Foss. 1970. "Decision processes during sentence comprehension: effects of surface structure reconsidered." Perception and Psychophysics. 8.413-416.

- Halle M., J. Bresnen and G. A. Miller. 1978. Linguistic Theory and Psychological Reality. MIT Press.
- Harmer L. C. and F. Norton. 1957. A Manual of Modern Spanish. University Tutorial Press. Second Edition.
- Harris M. 1976. Romance Syntax: Synchronic and Diachronic Perspectives. University of Salford.
- Harrison G., A. Triesman and N. Moray n.d. "The assessment of information content by Cloze Procedure." unpublished manuscript.
- Haworth S. 1979. Spanish Word Association Norms for 100 words of the Kent-Rosanoff list. M.A. thesis, Birkbeck College, London.
- Hayhurst H. M. 1969. Some Aspects of Children's linguistic Behaviour related to Memory. Ph.D. Thesis. Cambridge University.
- Horton D. L. and J. J. Jenkins. 1971. The Perception of Language. Merrill. Columbus Ohio.
- Hughes A. 1979. "Aspects of a Spanish Adult's Acquisition of English." Interlanguage Studies Bulletin. Utrecht. 4.1.49-65.
- Johnson N. F. 1965. "The Psychological Reality of Phrase-structure Rules." Journal of Verbal Learning and Verbal Behaviour. 4.469-475.
- Johnson N. F. 1968. "Sequential Verbal Behaviour." in Dixon T. R. and D. L. Horton (1968) (eds.)
- Juilland A. and E. Chang-Rodriguez. 1964. Frequency Dictionary of Spanish Words. Mouton.
- Kaplan R. M. 1972. "Augmented Transition Networks as Psychological Models of Sentence Comprehension." Artificial Intelligence. 3.77-100.

- Horton D. L. and J. J. Jenkins. 1971. The Perception of Language.  
Merrill. Columbus Ohio.
- Hughes A. 1979. "Aspects of a Spanish Adult's Acquisition of English."  
Interlanguage Studies Bulletin. Utrecht. 4.1.49-65.
- Johnson N. F. 1965. "The Psychological Reality of Phrase-structure  
Rules." Journal of Verbal Learning and Verbal Behaviour. 4.469-475.
- Johnson N. F. 1968. "Sequential Verbal Behaviour." in Dixon T. R. and  
Horton D. L. (1968) (eds.)
- Juilland A. and E. Chang-Rodriguez. 1964. Frequency Dictionary of  
Spanish Words. Mouton.
- Kaplan R. M. 1972. "Augmented Transition Networks as Psychological  
Modals of Sentence Comprehension." Artificial Intelligence.  
3.77-100.
- Kempen G. 1976. "Syntactic constructions as retrieval plans."  
British Journal of Psychology. 67.2-149-160.
- Kennedy R. A. 1968. "Response latency in the serial learning of  
short lists." British Journal of Psychology. 59.1-5
- Kennedy R. and A. Wilkes. 1968. Response times at different positions  
with a sentence. Quarterly Journal of Experimental Psychology.  
20.390-294.
- Kennedy R. and A. Wilkes. 1969. "An analysis of storage and retrieval  
processes in memorizing simple sentences." Journal of Experimental  
Psychology. 80.396-398.
- Kennedy R. and A. Wilkes. 1970. "A comparison of retrieval and  
location operations in active and passive sentences." British  
Journal of Psychology. 61.4.491-198.

- Kempen, G. 1976. "Syntactic constructions as retrieval plans." British Journal of Psychology. 67.2.149-160.
- Kennedy R. A. 1968. "Response latency in the serial learning of short lists." British Journal of Psychology. 59.1-5.
- Kennedy R. and A. Wilkes. 1968. Response times at different positions with a sentence. Quarterly Journal of Experimental Psychology. 20.390-394.
- Kennedy R. and A. Wilkes. 1969. "An analysis of storage and retrieval processes in memorizing simple sentences." Journal of Experimental Psychology. 80.396-398.
- Kennedy R. and A. Wilkes. 1970. "A comparison of retrieval and location operations in active and passive sentences." British Journal of Psychology. 61.4.491-498.
- Kershner J. and A. Jeng. 1972. "Dual functional hemispheric asymmetry in visual perception: effects of ocular dominance and post-exposural processes." Neuropsychologia. 10.437-445.
- Kimball J. 1973. "Seven Principles of Surface Structure parsing in natural languages." Cognition. 2.1
- Kinsella V. (ed.) 1978. Language Teaching and Linguistics: Surveys. Cambridge University Press.
- Kornfeld J. R. 1972. "Syntactic structure and the perception of sentences: some evidence of dominance effects." Proceedings of the 9th Annual Meeting of the Chicago Linguistics Society.
- Kornfeld J. R. 1973. "Clause Structure and the perceptual analysis of sentences." MIT Quarterly Report. 108.277-279.
- Kuhn T. S. 1962. The Structure of Scientific Revolutions. Chicago University Press.

- Ladefoged P. and D. E. Broadbent. 1960. "Perception of sequence in auditory events." Quarterly Journal of Experimental Psychology. 13.162-170.
- Lado R. 1965. "Memory Span as a factor in second language learning." International Review of Applied Linguistics. 3.2.123-129.
- Lambert W. E. 1955. "Measurement of the linguistic dominance of bilinguals." Journal of Abnormal and Social Psychology. 50.197-200.
- Lambert W. E., J. Havelka and R. Gardner. 1959. Linguistic manifestations of bilingualism." American Journal of Psychology. 72.77-82.
- Lambert W. E., M. Ignatow and M. Krauthamer. 1968. "Bilingual Organization in Free Recall." Journal of Verbal Learning and Verbal Behaviour. 7.204-214.
- Lawson E. 1961. "A note on the influence of different orders of approximation to the English language upon eye-voice span." Quarterly Journal of Experimental Psychology. 13.53-55.
- Lawson J., A. McGhie and J. Chapman. 1964. "Perception of speech in schizophrenia." British Journal of Psychiatry. 110.375-380.
- Leeson R. 1975. Fluency and Language Teaching. Longmans.
- Levelt E. J. M. 1970. "A scaling approach to the study of syntactic relations." in Flores d'Arcais and Levelt (eds.) 1970.
- Levelt W. J. M. 1974. Formal Grammars in Linguistics and Psycholinguistics. Three Vols. Mouton.
- Levelt W. J. M. and G. B. Flores d'Arcais. 1979. Studies in the Perception of Language. Wiley.

- Levy R. and A. Maxwell. 1968. "The effect of verbal context on the recall of schizophrenic and other psychiatric patients." British Journal of Psychiatry. 114.311-316.
- Long J. and E. Harding-Esch. 1977. "Summary and Recall of texts in first and second languages: some factors contributing to performance differences." in Sinaiko and Gerver (eds.) 1977.
- Luce R., R. Bush and E. Galanter. 1963. Handbook of Mathematical Psychology. volIII. Wiley.
- Lukatela G., M. Savic., P. Ognjenovic and M. Turvey. 1978. "On the relationship between processing the Roman and the Cyrillic Alphabets - a preliminary analysis with bialphabetic readers." Language and Speech. 21.113-142.
- Macnamara J. 1967. "The effects of instruction in a weaker language." Journal of Social Issues. 23.2120-134.
- Macnamara J. 1970. "Comparative studies of Reading and problem solving in two languages." TESOL Quarterly. 4.2.107-116.
- Macnamara J., M. Feltin, M. Hew and M. Klein. 1968. "An analytic comparison of reading in two languages." Irish Journal of Education. 2.41-53.
- Maher B. 1972. "The language of schizophrenia: a review and interpretation." British Journal of Psychiatry. 120.3-17 .
- Maher B. 1977. Contributions to the Psychopathology of Schizophrenia. Academic Press.
- Maidment E. 1973. Perceptual Organization in second language learning. M.A. Thesis. Birkbeck College, London.
- Marks M. R. and O. Jack. 1952. "Verbal context and memory span for meaningful materials." American Journal of Psychology. 65.298-300.



Meara P.M. 1978. "Learners' Word Associations in French."  
Interlanguage Studies Bulletin. 3.2.192-211

- Martin E. 1970. "Towards an analysis of subjective phrase structure." Psychological Bulletin. 74.153-166.
- Martin E. and K. Roberts. 1966. "Grammatical factors in sentence retention." Journal of Verbal Learning and Verbal Behaviour. 5.211-218.
- Matthews P. H. 1974. "Morphology: an introduction to the theory of word-structure." Cambridge University Press.
- McNeill D. 1970. The acquisition of Language: the study of developmental Psycholinguistics. Harper-Row.
- Meara P. M. 1975. "A note on Statistical Approximations." York Papers in Linguistics. 5.149-160.
- Meara P. M. 1978. "Schizophrenic Symptoms in Foreign Language Learners." UEA Papers in Linguistics. 7.22-49.
- Miller G. A. 1956. "The magical number seven plus or minus two+ some limits on our capacity to process information." Psychological Review." 56.81-97.
- Miller G. A. 1962. "Some psychological studies of grammar." American Psychologist. 17.748-762.
- Miller G. A. and S. Isard. 1963. "Some perceptual consequences of linguistic rules." Journal of Verbal Learning and Verbal Behaviour. 2.217-228.
- Miller G. A. and J. A. Selfridge. 1950. "Verbal Context and the recall of meaningful material." American Journal of Psychology. 63.176-185.
- Mills A. E. 1977. First and second language acquisition in German: a parallel study. Ludwigsburg.

Ochsner R. 1979. "A Poetics of Second Language Acquisition."  
Language Learning. 29.1.105-121

---

- Mishkin M. and D. Forgay. 1952. "Word Recognition as a function of retinal focus." Journal of Experimental Psychology. 43.43-48.
- Morris S. 1977. The effect of structural cohesion on visual perception: a comparison of native speakers and learners. M. A. thesis. Birkbeck College, London.
- Morris B. 1978. "Syntactic Structure and Visual Perception in Foreign Language Learners." Polyglot.1.1.D7-E1.
- Morton J. 1964. "The effects of context upon speed of reading, eye-movements and eye-voice span." Quarterly Journal of Experimental Psychology. 16.4.340-354.
- Morton J. 1978. "Word Recognition." in Morton J. and J. Marshall (eds.) 1978.
- Morton J. and J. Marshall. (eds) 1978. Psycholinguistics Series 2. Elek Science.
- Neale J. and R. Cromwell. 1977. "Attention and Schizophrenia." in Maher B. (ed.) 1977.
- Neisser U. 1967. Cognitive Psychology. Appleton Century Croft.
- Oldfield R. C. and J. C. Marshall. 1968. Language. Penguin.
- Oller J. W. 1972. "Scoring methods and difficulty levels for Cloze Tests of E.S.L. proficiency." Modern Language Journal. 56.151-158.
- Oller J. W. 1973. "Cloze Tests of second language proficiency and what they measure." Language Learning. 23.1.105-118.
- Oller J. W., J. D. Bowen, Ton That Dien and V. W. Mason. 1972 "Cloze Tests in English, Thai and Vietnamese: native and non-native performance." Language Learning. 22.1-15.

Peters A.M. "The Units of Language Acquisition." Working Papers  
in Linguistics. Hawaii. 12.1.1-72

Popper K. 1965. The Logic of Scientific Discovery. Harper Row.

- Oller J. W. and C. Conrad. 1971. "The Cloze Test and E.S.L. Proficiency." Language Learning. 21.183-196.
- Orbach J. 1953. "Visual fields as a function of cerebral dominance and reading habits." Neuropsychologia. 5.127-134.
- Orbach J. 1953. "Retinal locus as a factor in the recognition of visually perceived words." American Journal of Psychology. 65.555-562.
- Pavy D. 1968. "Verbal Behaviour in Schizophrenia: a review of recent studies." Psychological Bulletin. 70.164-178.
- Peranteau P., J. Levi and G. Pnares. 1972. The Chicago Which Hunts: Papers from the Relative Clause Festival. Chicago Linguistics Circle.
- Pimsleur P. 1966. Language Aptitude Battery. Harcourt Brace and World.
- Pimsleur P. and T. Quinn. (eds) 1971. The Psychology of Second Language Learning. Cambridge University Press.
- Postal P. M. 1967. "On so-called 'Pronouns' in English." Georgetown Monograph Series in Language and Linguistics. 19.177-206.
- Postman L. and Adams. 1960. "Studies in incidental learning VII. The effects of contextual determination." Journal of Experimental Psychology. 59.153-164.
- Postman L. and G. Keppel. 1970. Norms of Word Association. Academic Press.
- Raugh M. R. and R. C. Atkinson. 1975. "A mnemonic method for learning second language vocabulary." Journal of Educational Psychology. 67.1.1-16.

Rosansky E.J. 1976. "Methods and Morphemes in Second Language Acquisition Research". Language Learning. 29.409-425.

Seitz M.R. and S. Springer-Bloom. 1977. "Changes in Click Migration as a function of Age." Human Communication. 2.119-130

- Reed H. B. 1924. "Repétition and association in learning." Pedagogical Seminary. 31.147-155.
- Richards J. C. 1974. Error Analysis: Perspectives on Second Language Acquisition. Longmans.
- Richardson P. and J. Voss. 1960. "Replication Report; verbal context and the recall of meaningful material." Journal of Experimental Psychology. 60.417-418.
- Rosenzweig M. R. 1969. "International Kent-Rosanoff word association norms, emphasizing those of French male and female students and French workmen." in Postman and Keppel (eds.) 1969.
- Salzinger K., S. Portnoy and R. S. Feldman. 1962. "The effect of order of approximation to the statistical structure of English on the emission of verbal responses." Journal of Experimental Psychology. 64.52-57.
- Salzinger K., S. Portnoy and R. S. Feldman. 1964. "Verbal behaviour of schizophrenics and normal subjects." Annals of the New York Academy of Science.
- Schachter J. 1974. "An error in error analysis." Language Learning. 24.2.205-215.
- Shannon C. E. 1948. "A mathematical theory of communication." Bell System Technical Journal. 27.374-423. and 623-556.
- Sharp H. C. 1958. "Effect of contextual constraint upon recall of verbal passages." American Journal of Psychology. 71.568-572.
- Simon H. A. 1974. "How big is a chunk?" Science. 183.482-488.
- Sinaiko H. W. and D. Gerver. (eds.) 1979. Proceedings of the NATO Symposium on Language, Interpretation and Communication. Plenum.



Smith N.V. 1975. Review of Derwing 1973. Journal of Linguistics.  
11.2.261-270

- Slobin D. 1966. "Grammatical Transformations and sentence comprehension in childhood and adulthood." Journal of Verbal Learning and Verbal Behaviour. 5.219-227
- Slobin D. 1970. Universals of grammatical development in children. in Flores d'Arcais and Levelt (eds.) 1970.
- Smith S. 1954. Paper presented to the Eastern Psychological Association. cited in Miller 1955.
- Smith J. E. K. 1976. "Data transformations in Analysis of Variance." Journal of Verbal Learning and Verbal Behaviour. 15.3.339-346.
- Sperling G. 1960. "The information available in brief visual displays." Psychological Monographs. 74. no 498.
- Spolsky B., M. Sato, E. Walker and C. Arterburn. "Preliminary studies in the development of techniques for testing overall second language proficiency." Language Learning. Special Issue. 3.79-101.
- Sternberg S. 1967. "Retrieval of contextual information from memory." Psychonomic Science. 8.55-56.
- Stockwell R., J. D. Bowen and S. Martin. 1965. The Grammatical Structures of English and Spanish. University of Chicago Press.
- Stockwell R., P. Schachter and B. Hall-Partee. 1973. The Major Syntactic Structures of English. Holt Rinehart and Winston.
- Suci G., P. Ammon and P. Gamlin. 1967. "The validity of the probe latency technique for assessing structure in language." Language and Speech. 10.69-80.
- Traft M. and K. I. Forster. 1975. "Lexical storage and retrieval of prefixed words." Journal of Verbal Learning and Verbal Behaviour. 14.638-647.

- Taylor A. and N. Moray. 1960. "Statistical approximations to English and French." Language and Speech. 3.7-10.
- Taylor W. L. 1953. "Cloze procedure: a new tool for measuring readability." Journalism Quarterly. 3.415-433.
- Triesman A. M. 1965. "Verbal responses and contextual constraints in language." Journal of Verbal Learning and Verbal Behaviour. 4.118-128.
- Triesman A. M. 1965. "The effects of redundancy and familiarity on translating and repeating back a foreign and a native language." British Journal of Psychology. 56.369-379.
- Trim J. L. M. 1970. unpublished report: personal communication.
- Truscott T. P. 1970. "Contextual constraint and schizophrenic language." Journal of Consulting and Clinical Psychology. 35.189-194.
- Tulving E. and J. Petkau. 1962. "Concurrent effects of contextual constraint and word frequency on immediate recall and learning of verbal material." Canadian Journal of Psychology. 16.83-95
- Twomey E. 1979. A bibliography of research carried out in the field of vocabulary Learning in Second Languages. M.A. thesis, Birkbeck College, London.
- Valdman A. 1975. Bibliography of language learners' approximative systems and error analysis. Washington. Centre for Applied Linguistics.
- Vernon M. D. 1971. Reading and its difficulties. Cambridge University Press.
- Wales R. J. and E. Walker. 1976. New Approaches to Language Mechanisms. North-Holland

- Walters J. and R. Zatorre. 1978. "Latency differences for word recognition in bilinguals." Brain and Language. 6.158-167.
- Wanner E. and Maratsos M. 1976. "An ATN approach to comprehension." in Halle, Bresnan and Miller (eds.) 1978.
- Wilkes A. L. and R. Kennedy. 1969. "Relationship between parsing a and retrieval latency in sentences of varying grammatical form." Journal of Experimental Psychology. 79.241-245.
- Wilkes A. L. and R. Kennedy. 1970. "Response retrieval in active and passive sentences." Quarterly Journal of Experimental Psychology. 22.1-8.
- Wilkins D. A. 1972. Linguistics in Language Teaching. Arnold.
- Woodworth R. S. 1938. Experimental Psychology. New York.
- Yngve V. H. A. 1960. "A model and an hypothesis for language structure." Proceedings of the American Philological Society. 444-466.
- Zipf G. K. 1949. Human Behaviour and the Principle of Least Effort. Addison-Wesley.