palgrave macmillan

Understanding Complex Sentences

Native Speaker Variation in Syntactic Competence

Ngoni Chipere

Understanding Complex Sentences

Understanding Complex Sentences

Native Speaker Variation in Syntactic Competence

Ngoni Chipere

palgrave
macmillan

First published 2003 by
PALGRAVE MACMILLAN
Houndmills, Basingstoke, Hampshire RG21 6XS and
175 Fifth Avenue, New York, N.Y. 10010
Companies and representatives throughout the world

PALGRAVE MACMILLAN is the global academic imprint of the Palgrave Macmillan division of St. Martin's Press, LLC and of Palgrave Macmillan Ltd. Macmillan® is a registered trademark in the United States, United Kingdom and other countries. Palgrave is a registered trademark in the European Union and other countries.

ISBN 0–333–98639–3 hard back

This book is printed on paper suitable for recycling and made from fully managed and sustained forest sources.

A catalogue record for this book is available from the British Library.

Library of Congress Cataloging-in-Publication Data

Chipere, Ngoni, 1965–
Understanding complex sentences : native speaker variation in syntactic competence / Ngoni Chipere.
p. cm.
Includes bibliographical references and index.
ISBN 0–333–98639–3
1. Grammar, Comparative and general—Syntax.
2. Psycholinguistics. 3. Language and languages—Variation.
I. Title.

P295.C485 2003
415—dc21

2003042939

10 9 8 7 6 5 4 3 2 1
12 11 10 09 08 07 06 05 04 03

Printed and bound in Great Britain by
Antony Rowe Ltd, Chippenham and Eastbourne

I dedicate this book to my parents – Benjamin Nyikadzino and Mary Virginia Chipere.

Contents

List of Figures

List of Tables

Preface

When I told my MA supervisor, Ewa Dabrowska, that I was thinking of studying native speaker variations in syntactic competence, she warned me, from personal experience, that this would not be a clever career move. She said that I would find it difficult to get my work published. Her warning got me hooked instantly. Things that people don't like to think about usually turn out to be the most intellectually challenging and the most rewarding.

This has certainly proved to be the case with native speaker variations. This topic connects a surprisingly wide range of issues in linguistics and psychology. Tracing these connections provides an opportunity to break a stalemate which has stymied theoretical development in various subfields of linguistic and psychological research over the past fifty years. These connections reveal that language and memory are organised as a network. This conclusion suggests that linguistics and psychology can interact more productively by making use of a common mathematical language – graph theory – in order to describe linguistic structures and processes. The network structure of language and memory also indicates that human mental faculties are much more powerful than has so far been supposed. There are some highly positive implications of this for language education and education in general.

I provide a synopsis of the book in the first chapter for the reader's convenience. The book covers developments spanning a century and covering more than one discipline. It is therefore difficult to avoid certain terminological problems, for which I beg the reader's indulgence.

Acknowledgements

I would like to thank John Williams, for supervision of the thesis on which the book is based and Ewa Dabrowska for first introducing me to the topic. Thanks also to Pilar Duran, Brian Richards, David Malvern and Debra Powell for reading and commenting on various chapters. Many thanks to Jill Lake of Palgrave Macmillan and an anonymous reviewer. I am very grateful to Richard Hudson, for encouragement and constructive criticism. The shortcomings of the book are, of course, entirely my own. Finally, but certainly not least, many thanks to Luli and to my sisters Grace, Becky and Rumbi for their support during the PhD ordeal.

1
Introduction

The most complex inventions of syntacticians are sometimes no match for sentences that occur naturally. Numerous examples of the following kind can be found in the British National Corpus:

> Consequently, the answer to be given to the national court must be that the fact that the competent minister of a member state has the power to dispense with the nationality requirement in respect of an individual in view of the length of time such individual has resided in that member state and has been involved in the fishing industry of that member state cannot justify, in regard to Community law, the rule under which registration of a fishing vessel is subject to a nationality requirement and a requirement as to residence and domicile (*The Weekly Law Reports*, 1992 in the British National Corpus).

Legalese is, of course, notoriously convoluted. Nevertheless, its convolutions lie completely within the syntactic possibilities of the English language. Judging from the many examples found in the British National Corpus and other sources, syntactic complexity is not confined to the legal domain. Here is another, less complex example:

> THAT variation in word order does indeed indicate different information structure can be seen from the 'fact' that, even though different variations might be equally well-formed, they are not necessarily equally interchangeable in a given context (Kruijff, 2002, writing to the electronic discussion group *Corpora*).

Though such examples may predominate in certain kinds of discourse, their existence shows that highly complex sentences are part of the

English language. This fact raises the following problem. It is widely assumed that speakers have a uniform and automatic ability to comprehend all possible grammatical sentences in their native language. This assumption underlies much of the purely theoretical research in linguistics and a considerable proportion of the experimental work in psycholinguistics. It appears, however, that the assumption is wrong. There are many native speakers of English who cannot understand even moderately complex sentences, such as the following:

> The doctor knows that the fact that taking care of himself is necessary surprises Tom.

In comparison to the authentic sentences presented above, this artificial experimental stimulus would seem like child's play. Yet, Dabrowska (1997) found that there are native English speakers who simply cannot understand such sentences. When she asked participants in her experiment to answer simple comprehension questions such as, 'What does the doctor know?' or 'What surprises Tom?', she found that university lecturers performed better than undergraduates, who, in turn, performed better than cleaners and porters, most of whom completely failed to answer the questions correctly.

I replicated Dabrowska's key results and witnessed first-hand the genuine difficulties experienced by otherwise normal, intelligent and articulate native English speakers as they grappled with the unfamiliar syntax. Their difficulties were not attributable to performance factors such as memory limitations or the conceptual content of the sentences. It appeared that they had not encountered such sentences before and simply did not know how to decode them. Many complained that there were too many 'thats' in the sentence and they thought the sentence might read better if the second 'that' were removed. Of course, that would only render the sentence ungrammatical. The participants had read the sentences using a computer program which recorded the amount of time it took them to read each word. The reading time data showed that participants took considerably longer to read the second 'that'. In another experiment, which involved getting participants to recall the test sentences, the most frequently forgotten word was the second 'that'. It seemed as though participants were not expecting to encounter two complementisers in rapid succession and were thrown completely out of kilter by the unfamiliar sequence. This would suggest that they were not analysing the syntax directly but processing the sentences by analogy to a simpler and more familiar type of sentence, something perhaps along the lines of,

The doctor knows that taking care of himself is necessary.

It is interesting to note that, in the experiment involving recall, many of the words in the test sentences which did not conform to this simpler pattern were simply forgotten. This observation echoes a previous finding by Labov (1979), who found that some individuals found it impossible to accurately repeat sentences whose grammatical construction did not conform to their dialect of English.

In another experiment, I also found, unexpectedly, that post-graduate students who were *not* native speakers of English performed better than native English cleaners and porters and, in some respects, even performed better than native English post-graduates. Presumably this was because non-native speakers actually learn the grammatical rules of English, whereas explicit grammatical instruction has not been considered necessary for native speakers, at least until quite recently in England.

At this point, I should make it clear that individual differences in comprehension are not confined to speakers of English. It is important to emphasise this point because I have discovered that my findings are prone to a particularly unfortunate interpretation. A number of reviewers have mistakenly assumed that the primary intention of my experiments was to poke fun at native English speakers by suggesting that some of them cannot speak their own language well. As the reader will soon make out, however, my interest in individual differences is somewhat more abstruse than that. In any case, the review of the experimental literature in Chapters 3 and 8 will show that individual differences have been reported not only among native speakers of English, but also among native speakers of Somali, Italian, Spanish and Japanese. These differences range over knowledge of phonology, morphology, vocabulary, syntax and text structure.

This evidence indicates that linguistic abilities are not distributed uniformly across a population of native speakers. This conclusion poses such serious theoretical and ideological problems that many linguistic researchers refuse to even think about it. From a theoretical point of view, individual differences threaten the conceptual foundations of much linguistic and psycholinguistic thought. From an ideological point of view, researchers object to the idea that some individuals can use language better than others. The effect of these two attitudes has been to suppress evidence of individual differences for more than forty years. Newcomers to the field are unaware of this evidence because it is rarely cited and they tend to adopt and perpetuate the conventional, though unproven, assumption that native speakers have a uniform grammatical competence. It is interesting to note that researchers who somehow discover individual differences in syntactic competence invariably express great surprise at this discovery, in spite of the fact that evidence for individual differences goes back at least as early as Bloomfield (1927). Clearly, it is necessary to address the two attitudes to individual differences squarely before any real progress can be made on this issue.

Individual differences disconcert particularly generative grammarians and generative psycholinguists because it conflicts directly with their central axiom, which is that speakers have a uniform competence in the grammar of their native language. In personal communication, many have claimed that no such assumption exists. The truth, however, is that much of generative linguistics and psycholinguistics stands or falls with this assumption as can be inferred from the following observations. Firstly, generative linguists often use their personal native speaker intuitions as primary data. This methodological practice is unjustifiable in the absence of the assumption that other native speakers share those intuitions. I should point out that this practice is not confined to generative linguistics but to many other theoretical approaches in linguistics which do not employ empirical methods. Secondly, experimental psycholinguists routinely use the performance of university student subjects in order to make inferences about the linguistic abilities of *all* native speakers of a language. The use of such a highly unrepresentative sample can only be justified in terms of a prior assumption that native speakers have a uniform grammatical competence. These two observations indicate that, were the assumption of uniform competence ever to be relaxed, then many theoretical and experimental results would immediately become suspect. It is therefore not too surprising that evidence of individual differences is routinely ignored.

In numerous personal exchanges, I have also found that evidence of individual differences disturbs some linguists for ideological rather than theoretical concerns. For these linguists, it sounds distinctly and dangerously politically incorrect to suggest that less educated people are less syntactically competent than more educated people. It sounds rather like a step back in time to the days when working class Britons and African–Americans were thought to speak defective variants of English. This fear is based on a misapprehension, however. Individual differences in understanding complex sentences are not related exclusively to level of education. As I will show in Chapter 8, pre-school children differ significantly in linguistic ability. I also describe an experiment in Chapter 9 which produced evidence of individual differences among native English speakers who were matched for age and the number of years spent in formal education. There are also highly educated native English speakers who were unable to understand complex sentences in an experiment described in Chapter 9.

As things currently stand, a strong mixture of ideological and theoretical objections has created a powerful taboo on the subject of individual differences. As a result, reliable experimental evidence of such differences has been effectively ignored for the last forty years. This refusal to confront uncomfortable facts creates both theoretical and moral problems. I describe the theoretical problem first and then discuss its social ramifications.

In numerous experiments, measurements of sentence processing such as comprehension, recall and reading time profiles have all provided reliable evidence that native speakers differ in linguistic performance (see, for instance, Caplan and Waters, 1999; Just and Carpenter, 1992; MacDonald, 1994). The accumulated evidence creates the following dilemma for anyone wishing to maintain the assumption of uniform native speaker competence. In order to explain individual differences in performance, it is necessary to locate the source of individual inequality elsewhere. Some psycholinguists have done this by invoking a distinction between competence and performance. They have suggested that native speakers share the same grammatical competence but differ in a performance factor called 'working memory'.

There is no universal agreement on just what working memory is, but it can be thought of as mental processing capacity, something that determines the speed of mental operations and the amount of information which can be processed at any one time (see Caplan and Waters, 1999; Just and Carpenter, 1992 for two different definitions of working memory). Individuals with high working-memory capacity are said to understand complex sentences better than individuals with low working-memory capacity. On the face of it, this solution appears to save the assumption of uniform native speaker competence in a rather neat fashion. However, it does so in a highly implausible fashion. Working memory is said to be an individual trait, something that is innate to the individual and not subject to alteration. Individuals with low working-memory capacity are therefore, in a sense, mentally deficient, given that they lack enough of a mental property which is vital for mental performance. In the final analysis, therefore, the working-memory hypothesis saves uniform native speaker competence by classifying large sections of the human population as incurably mentally deficient.

I do not wish to suggest that psycholinguists have been motivated by ideological concerns in advancing the working-memory hypothesis, only that they have sought an explanation of individual differences which maintains the assumption of uniform native speaker competence. The point is that anyone wishing to maintain that assumption for ideological reasons will find themselves in exactly the same situation as the psycholinguists who have advanced the working-memory hypothesis. They would have to posit inequality elsewhere in order to explain individual differences. In ideological terms, the end result would be the same, namely that some individuals can use their native language better than others.

One way out of the dilemma might be to suggest that speakers have different mental grammars of their native language and that some mental grammars do not generate complex structures such as the examples given earlier. This proposal has a surface appeal in that it suggests that individual grammars are not unequal, simply different. However, this proposal reduces ultimately to the view that there are native speakers who cannot understand complex sentences simply because they lack the grammatical knowledge required to do so. Ideological objections to the study of individual differences are therefore neither here nor there because the bare fact of individual variations remains to be explained.

Whether it is driven by purely theoretical or purely ideological concerns, the assumption of uniform native speaker competence has a pernicious social side effect. Linguistic competence is a prerequisite for academic development and schoolchildren with insufficient knowledge of their native language are severely disadvantaged. The failure to confront all the facts concerning individual differences means that the wrong message is being given to educationalists. As I described earlier, the inability to understand complex sentences is conventionally attributed to a shortage of working-memory capacity, which is effectively a mental disability. This view has negative educational implications because it is thought that nothing can be done to increase working-memory capacity other than to teach victims strategies that help them circumvent their working-memory deficiencies (see Oakhill, 1994; Perfetti, 1994). However, I found that, if given appropriate instruction, poor comprehenders can attain high levels of sentence comprehension and recall (see Chapter 9). It therefore appears that there is nothing inherently defective about these individuals: they simply do not have the grammatical knowledge needed to understand complex sentences.

Given that this knowledge can be taught, the implications for education, which I discuss in Chapter 10, are therefore positive. In order to appreciate these implications fully, it is necessary to put ideological concerns aside and consider the issues objectively. As it turns out, the objective approach that I adopt in this book actually gives rise to a deeply egalitarian view of the human potential for language.

In this book, I seek to decide between two psychological explanations of native speaker variations. One explanation attributes them to individual variations in working-memory capacity and another to individual variations in linguistic experience. These alternative explanations arise from two theoretical traditions which I refer to in this book as rule-based and experience-based approaches. There may, of course, be other possible explanations but I restrict my attention to these two because they are highly general. They are not confined to the narrow issues of individual differences and complex sentences but deal with a wide range of issues about the nature of linguistic knowledge and the psychological mechanisms underlying language use. Individual differences represent a minor and marginal case in point for each approach. However, it is a case in point which happens to touch most directly on the central assumptions of each approach. The explanations which each approach offers to account for individual differences arise from deep theoretical principles about language and the psychology of language. Consequently, discoveries concerning individual differences have logical implications for the central assumptions of each approach as well as the wide range of issues that are treated by each tradition. This book is ultimately concerned with these general issues.

1.1 Rule-based and experience-based approaches

The differences between rule-based and experience-based approaches can be illustrated conveniently in terms of an analogy to two individuals with different kinds of mathematical knowledge. One individual has learned the times tables purely by rote and without any knowledge of the principles of multiplication. The other individual has learned the rules of multiplication but has not memorised any times tables. These two individuals will perform differently on a range of multiplication problems on account of the qualitative differences in their knowledge of multiplication. The rote-learner will perform flawlessly on learned problems but will fail to generalise his or her rote-learned knowledge to novel problems. The rule-follower will be able to solve both familiar and novel problems excepting those that are so complex that they overburden his or her short-term memory (STM).

The rule-based approach describes language users as rule-followers who possess grammatical rules of sufficient descriptive power to enable the comprehension of any sentence in their native language. This approach also makes a fundamental distinction between competence and performance. It is thought that while language users possess the grammatical mastery to understand all the possible sentences of their native language, limits in working memory make it impossible to understand sentences whose complexity exceeds a certain level. The competence–performance distinction is made to account for individual differences in the following way. It is said that speakers possess a uniform competence in the grammar of their native language but vary in working-memory capacity. As a result, comprehension failure is said to occur at levels of syntactic complexity that vary across individuals (see Just and Carpenter, 1992). It is notable that proponents of the rule-based approach evince great interest in working-memory processes, which are involved in symbol manipulation, almost to the exclusion of long-term memory (LTM) processes, which are involved in learning.

The experience-based approach likens language users to rote-learners, whose capacious LTM allows them to learn all the types of utterances that are *likely* to occur in their language. It is important to note, however, that this approach also attributes to language users a powerful psychological mechanism of generalisation which enables them to understand sentences that are similar to but not necessarily identical to those encountered in the past. It is important to emphasise this last point because the experience-based approach has been unfairly caricatured in the past as implying that sentences can only be understood if they are exact replicas of those stored in memory.

The experience-based approach explains individual differences in terms of variations in experience. This approach does not make a distinction between competence and performance. It does, however, impose a limit on the human ability to understand complex sentences. This limit arises naturally from the way this approach represents relationships between linguistic units in a sentence. In contrast to the rule-based approach, which represents syntactic relationships in terms of hierarchical phrase structure configurations, the experience-based approach represents syntactic relationships in terms of associations between units. This approach makes it possible to capture important statistical dependencies between units. However, it has a major disadvantage in that it cannot represent structured relationships that are arbitrarily complex. For instance, the associative networks used by proponents of the experience-based approach cannot represent the relationship between a matrix noun and a matrix verb when they are separated by an inordinately large amount of intervening material (see Elman, 1992). It is also notable that proponents of the experience-based approach demonstrate a great interest in LTM processes, which are involved in learning almost to the exclusion of working-memory processes, which are involved in symbol manipulation.

The opposition between rule-based and experience-based approaches runs deep. In linguistic theory, formal linguists disagree with various kinds of functional linguists on whether linguistic knowledge is an autonomous system of rules or a set of conventions that are shaped by contexts of use. In sentence comprehension studies, rule-based models of language processing stand in opposition to probabilistic models. In studies of language acquisition, the innateness hypothesis stands in opposition to the idea that knowledge of language is induced from experience. Finally, in language teaching, there is a long-standing and unresolved tension between rule-based and usage-based pedagogical approaches. But that is not all. The rift extends beyond language-related research into other fields.

If we step outside the domain of language studies and consider the wider field of cognitive science, we can see the same split. At this much more general level, we find an opposition between computationalism and connectionism. Computationalism views the human mind as a type of computing machine that operates according to rules encoded in its programming. Connectionism, on the other hand, views the mind as a network of associations that encode probabilistic patterns of co-occurrence between features of experience. If we widen our perspective even further, we can see that computationalism and connectionism are only instances of a much broader opposition in philosophy and psychology between rationalism and empiricism (or associationism). In Chapter 2, I will show how rationalist principles underlie the rule-based approach on the one hand, while associationist principles inform the experience-based approach on the other.

In the final analysis, therefore, rule-based and experience-based explanations of individual differences actually represent two ancient philosophical traditions concerning the nature of human knowledge and the operations of human mind. This is the depth to which the opposition between rule-based and experience-based approaches ultimately extends. While proponents of each approach may not necessarily be consciously aware of their philosophical commitments, the fact that they make these commitments can be inferred from their methodological practices and core assumptions. Chapter 2 will show that a wide range of the features that characterise rule-based and experience-based approaches in linguistics and psychology bear the hallmarks of rationalism and associationism. It is perhaps on account of such underlying and often unthought assumptions that the opposition between the two approaches has proven itself to be so intractable over the past half century.

As the following chapters will show, history has tended to repeat itself across the various sub-fields of linguistic research. One approach replaces the other in dominance cyclically. New generations of researchers think they are discovering new ways of doing things when they are simply repeating previous arguments. The cyclic pattern is encouraged by the need to refer to the most recent publications and a failure to take note of earlier research. As Chapters 2 and 4 will show, the repetitions can be quite striking, with almost identical statements being made by researchers separated by forty years. In brief, language-related research is locked in a stalemate.

I adopt two investigative strategies to break out of the impasse. Firstly, I carry out a literature review on the historical development of the two approaches in both linguistics and psychology. Secondly, I examine a narrow experimental question, to do with the causes of individual differences, whose answer impinges directly on the foundational assumptions of each approach. Both the literature review and the experimental work indicate that the two approaches provide partial and complementary explanations of individual differences as well as a range of other issues relating to language and its psychological representation. Below, I provide condensed descriptions of the theoretical and empirical arguments developed throughout the book in order to show that the two accounts are complementary.

1.2 The dual nature of linguistic cognition

From a purely logical standpoint, the partial correctness of each view can be seen by considering the main difference between them. Proponents of the rule-based approach have argued that language is a rule-governed phenomenon. Therefore, the argument has traditionally continued, knowledge of language must consist of rules. On the other hand, proponents of the experience-based view have argued that language has a statistical structure. Therefore, this alternative argument goes, knowledge of language must be based on experience. It is easy to show that these two seemingly opposed views are not mutually exclusive. This demonstration only requires us to consider the fact that highly regular domains such as mathematics have both rule-governed and experiential aspects and that mathematical reasoning is both rule-governed and memory-based.

Let us consider the role of formulae in mathematical reasoning. Mathematical formulae, such as the formula for calculating the area of a circle, are derived via often long and arduous processes of logical deduction. In general, once a useful mathematical expression has been obtained, it is applied on subsequent occasions in a ready-made form without repeating the process of its derivation. In a manner of speaking, formulae automate and accelerate mathematical reasoning by compressing long sequences of deduction into compact and mentally economical statements. If mathematicians did not record and reuse these results of previous deductions, they would be doomed to derive each formula afresh each time they needed to use it. Progress in mathematics would be impossible otherwise.

The point I wish to make here is that knowledge of mathematics is not purely deductive. A considerable portion of it consists of the repertoire of formulae, constants and techniques that individual mathematicians must learn by rote. Proponents of the rule-based approach often overlook the fact that mastery of a rule-governed domain such as mathematics requires years of practice and the accumulation of relevant experience. Consider, for instance, chess, which is another highly rule-governed domain. A mere knowledge of the rules of chess does not make one a grand master: it is only the minimum requirement for a mastery of chess. Among other things, extensive experience is also required, including memory for specific moves and games. It seems clear from all this that, *as a matter of principle*, knowledge of highly regular domains has both deductive and experiential dimensions. The two aspects are mutually dependent: the outputs of deduction are stored in memory and they subsequently form inputs into deductive processes.

The dual nature of knowledge in rule-governed domains provides a natural explanation of individual differences, for the following reason. The trained mathematician may know, at least in his or her own speciality, how to derive standard formulae and to generalise them whenever the need arises. However, non-mathematicians will have learned mathematical formulae by rote, without any knowledge of the preceding deductive processes. The piecemeal quality of such rote-learned knowledge means that non-mathematicians will find it difficult, if not practically impossible, to generalise it to novel problems. For instance, rote-learning the formula for calculating the area of a circle will not lead to an automatic ability to calculate the area of an ellipse, despite the fact that the expressions for calculating the two areas of the two geometric forms are almost identical. The dual nature of mathematical knowledge can therefore give rise to rule-followers and rote-learners.

These logical considerations can be extended to knowledge of language. If we assume, for the sake of argument, that knowledge of language is rule-based, then we need to consider the logical possibility that knowledge of language also has a dual nature and that native speakers of a language also divide into rule-followers and rote-learners. This possibility is strengthened by the fact that, in actuality, natural language is not as rule-governed as mathematics. If it *were* completely rule-governed, then grammarians would by now have constructed a purely rule-based grammar of at least one natural language. Their failure to do so, despite much effort over many years, indicates that the domain of natural language is considerably irregular (see Gross, 1979). This irregularity means that, whereas the memorised component of mathematical knowledge is ultimately consistent with a rule system, having been derived from it, the memorised component of linguistic knowledge is not completely consistent with any rule system. All speakers of a language are therefore, to some extent, necessarily rote-learners. With regard to the more systematic aspects of language, however, there are opportunities for native speakers to vary in terms of being rule-followers or rote-learners.

I have presented above a purely logical argument to show that human knowledge and cognition in highly regular domains have a dual character, incorporating rule-based and experiential aspects. This observation is not at all new. Computer scientists have long recognised a tension between computation and storage. If certain routine operations are performed repeatedly, is it more or less efficient to recalculate results every time or to store results and then simply look them up subsequently? There are situations when storage is the more efficient option. The question of computation versus storage has also been addressed with regard to the human use of language (see Nooteboom *et al.*, 2002; and also Wray, 2002). I will now present an empirical argument to show that native speakers have the dual potential for rule-governed and experience-based modes of comprehension.

Models of sentence comprehension have long been concerned with how language users process complex sentences, particularly *self-embedded sentences*. The existence of self-embedded sentences in natural language has been used by proponents of the rule-based approach to argue that knowledge of language is rule-governed. The argument is based on the fact that the ability to create and comprehend self-embedded sentences appears to be more consistent with a rule-based rather than a probabilistic knowledge of language. Self-embedded sentences are created by embedding one sentence within another. For instance, 'The dog is mad' can be expanded to 'The dog that the homeless person likes is mad'. This new sentence can be further elaborated as, 'The dog that the homeless person that the shoppers avoid likes is mad' and so on. It seems that the process of forming such sentences is based on a categorical rule which states that a noun phrase can be post-modified by a relative clause whether the noun phrase in question is in the matrix or in the relative clause. In principle, it is possible to carry out an arbitrary number of embeddings.

It has not been possible from the 1950s to date to create associative mechanisms which can support such a process to an arbitrary degree. The reason for this is as follows. Existing associative mechanisms cannot represent recursive rules (that is, rules that can be reapplied while still in the process of being applied). Instead, existing associative schemes only learn and represent sequences. This means that an associative mechanism can only represent self-embedded sentences as sequences. In addition, associative mechanisms cannot generalise much from shorter sequences to longer sequences. As a result, an associative mechanism can only represent self-embedded sentences of a length which is the same as that of sentences that it has been exposed to. In addition to this lack of generalisation, there is a further problem. There is a statistical limit to sequence length. Beyond a certain length, it becomes impossible to distinguish between two or more similar sequences. On account of these reasons, it is not possible, in terms of the experience-based approach, to produce or understand sentences that exceed two to three levels of self-embedding (see below; Christiansen and Chater, 1999).

A key difference between the two approaches is therefore as follows. The rule-based approach suggests the possibility of an infinite number of self-embeddings, subject to memory limitations but the experience-based approach places an upper limit of two or three degrees on self-embedding. The question then is, which of these two accounts is correct? Experimental evidence indicates that native English speakers have difficulty in understanding sentences containing two or more levels of self-embedding, though the precise degree of acceptable embedding varies in relation to semantic and other factors. The rule-based model of Gibson and Thomas (1999) explains the limit on self-embedding in terms of the competence–performance distinction. It is said that native speakers comprehend such sentences by using recursive rules to build syntactic trees in working memory. Comprehension is compromised if the syntactic tree exceeds working-memory capacity. On the other hand, the experience-based models of Elman (1992) and Christiansen and Chater (1999) explain the limit in terms of interference in associative networks. Self-embedding introduces material that intervenes between the matrix subject and the matrix predicate. This material creates interference which makes it difficult to distinguish between valid and invalid continuations of a word sequence. Comprehension fails when the level of interference exceeds a certain critical threshold.

It is notable that these two accounts restate, in much the same terms, two accounts of sentence comprehension proposed earlier by Chomsky (1957, 1965) and Hockett (1955). As I argue in Chapter 2, basic ideas about knowledge of language have remained unchanged over the past half century. In Chapter 2, I show that Hockett's (1955) finite state model and Elman's (1992) simple recurrent connectionist network are based on the idea that knowledge of language is an associative network which encodes probabilistic relationships between linguistic units. On the other hand, Chomsky's (1957, 1965) generative model and Gibson and Thomas' (1999) symbolic model are based on the idea that knowledge of language is a formal deductive system which is constrained by working-memory capacity. It is a measure of the depth of the stalemate between the rule-based and experience-based approaches that their disagreement over how the same type of sentence is processed has not altered for nearly half a century.

In Chapter 3, I will review a number of experimental studies carried out in the 1960s and 1970s that can shed light on the issue. These studies revealed two important facts that had the potential then, and still have now, to break the fifty-year impasse. It appears that sentence comprehension can make use of both LTM for linguistic experience and what may be described as formal rules. In the special case of self-embedded sentences, it appears that the reason why many native English speakers cannot understand sentences with two or more levels of self-embedding is not because they have insufficient working-memory capacity but simply because they have insufficient grammatical knowledge (see Blumenthal, 1966; Stolz, 1967). It appears that naïve experimental subjects process sentences on the basis of their prior linguistic experience and this experience often cannot be generalised to cover multiply self-embedded sentences. Secondly, however, if they are provided with grammatical instruction, native English speakers can acquire the ability to understand sentences with more than two degrees of self-embedding (see Blaubergs and Braine, 1974; Freedle and Craun, 1970; Stolz, 1967). The study by Blaubergs and Braine shows that, after training, STM constraints begin to operate at around five degrees of self-embedding.

In Chapter 4, I will show that the issue of self-embedded sentences continues to be debated at present in terms very much like those of Chomsky and Hockett. I will describe two current experience-based and rule-based models of sentence comprehension – Elman's (1992) connectionist model and Gibson and Thomas' (1999) symbolic model. I will argue that these models of sentence comprehension make assumptions about linguistic knowledge that are identical to those made by Chomsky and Hockett. I will also argue that current models fail to address many of the experimental findings reviewed in Chapter 3. In Chapter 5, I discuss three current theories of individual differences. MacDonald and Christiansen's (2002) model is based on the Elman (1992) model, while Just and Carpenter (1992) and Caplan and Waters' (1999) models make underlying assumptions similar to those of Gibson and Thomas (1999). I argue that these theories of individual differences suffer from limitations arising from their underlying principles and they are unable to account for the simultaneous existence of rule-governed and experience-based modes of sentence comprehension.

Neither approach, in its early or current form, therefore tells the whole story about individual differences. If anything, the two approaches paint *complementary* pictures of a richer reality. The experimental evidence suggests that language users are capable of comprehending sentences by analogy or by rule. As Chapter 3 will show, some native speakers appear to be rule-followers while others appear to be rote-learners. The simultaneous existence of rule-based and experience-based modes of sentence comprehension represents a major challenge for each approach, because it has always been assumed that the two possibilities are mutually exclusive. It appears, however, that the human mind is a resourceful system which blends rule-governed and experience-based modes in proportions that vary across individuals. Proponents of either approach have correctly identified one of these modes of operation but they have overgeneralised its scope and incorrectly assumed that it is the only one. This has left each approach unable to account for the alternative mode of comprehension.

Although I will confine the discussion to the distinction between rule-based and experience-based forms of linguistic knowledge, it is important to point out that the interaction of this pair of oppositions can lead to many kinds of individual differences. Firstly, individuals can vary in terms of being rule-followers or rote-learners (bearing in mind that this dichotomy marks two extremes of a continuum). Within each type, individuals might also differ. For instance, if linguists disagree amongst themselves concerning the grammar of a language, then there is no reason to assume that native speakers do not also vary. Therefore, it is possible that, at least in principle, rule-followers may differ amongst themselves in the rules that they use to process language. Rule-followers can also differ amongst themselves in the following way. There is psychological evidence which shows that experts develop mnemonic techniques to help them manage information required for the performance of certain tasks. An individual who has developed appropriate mnemonic techniques could therefore display higher levels of linguistic performance than one who has not, even if both individuals possess knowledge of the same grammatical rules.

Where no grammatical principles or memory management are involved, individuals might vary in their sensitivity to the statistical structure of language. I will review evidence in Chapters 3 and 5 which shows that individual differences in probabilistic knowledge have an impact on linguistic performance. It is also important to note that experiential and rule-based aspects of linguistic knowledge are not totally independent. Arguably, linguistic rules are induced from linguistic experience. Differences in experience can therefore lead to differences in knowledge of grammatical rules. An example of this pattern of individual differences is presented by Cupples and Holmes (1987, 1992), who show that differences in lexical knowledge can lead to differences in the ability to apply syntactic rules. Apparently, some native speakers of English cannot assign lexical items to their correct parts of speech. Given that syntactic rules operate on lexical categories, individual differences in lexical categorisation can be expected to lead to individual differences in the ability to apply syntactic rules, which is what Cupples and Holmes found. It is therefore possible, at least in principle, that two individuals who possess knowledge of the same grammatical rules might perform differently if they differ in lexical knowledge.

In theory, therefore, there may be a quite complex pattern variation. In addition to variations between the two groups, rule-followers and rote-learners, there can also be variations, as suggested above, within the groups themselves: rule-followers in terms of the actual rules they use, their lexical knowledge and their ability to manage working memory; rote-learners in their probabilistic knowledge as well as in lexical knowledge and working-memory management. All these variations can, however, be traced back to the distinction between rote-learners and rule-followers and I will limit the discussion to this distinction. As I have argued above, neither the rule-based nor the experience-based approach can explain this basic pattern of individual differences. In order to do so, it is necessary to have a general framework for the study of language and linguistic behaviour that integrates rule-based and experience-based perspectives.

A number of researchers recognise the need for an integrated approach and some of them have proposed hybrid models that combine aspects of computationalism and connectionism in the construction of models of sentence comprehension. In Chapter 5, I review two hybrid models of sentence comprehension, that of Miikkulainen (1996) and Townsend and Bever (2001). These models integrate aspects of rule-based and experience-based processing. I will argue that these two attempts are too narrowly based to have a general application. In view of the depth of the rift between rule-based and experience-based approaches, it is necessary to consider conceptual issues at a foundational level before proposing specific models of sentence comprehension.

I will therefore consider some basic issues in the main disciplines that feed into psycholinguistics – linguistics, psychology and mathematics. Psycholinguistics has tended to borrow from these disciplines in a superficial manner and the lack of conceptual depth has made it difficult to find solutions to the problem of integrating rule-based and experience-based approaches. One of the reasons why I take a historical approach in this book is to examine the contribution of these fields to psycholinguistics and to expose areas where ideas have been misinterpreted. Below, I describe the contribution of each field briefly in anticipation of more extensive discussion in later chapters. I will begin with linguistic issues.

Psycholinguists seem to be largely unaware of the fact that both rule-based and experience-based approaches in linguistics actually have a common source in the theory of Ferdinand Saussure. The rule-based approach focuses on that aspect of language which Saussure referred to as *la langue*, the abstract system of language, while the experience-based approach focuses on that aspect of language which Saussure referred to as *la parole*, what might be referred to in modern parlance as corpus data. Although Saussure's theory is quite old, having been proposed at the start of the last century, it shows, in embryonic form, how rule-based and experience-based perspectives can be integrated within a single linguistic framework and it does so in a psychologically plausible manner. By an interesting twist of history, the theory is actually based on the principles of 19th century associationist psychology. Current linguistic theories can therefore actually be traced back, at least in part, to psychology.

I will use Saussure's theory as a basis for integrating rule-based and experience-based approaches in linguistic terms.

Psycholinguists have also made psychological assumptions about the human memory system. However, they have done so without making adequate reference to the psychology of memory, as pointed out by Lewis (1996). This has led to a skewed approach to human memory. As I argue in Chapter 6, proponents of the rule-based approach have tended to focus on short-term or working-memory processes, while proponents of the experience-based approach have tended to focus on LTM processes. These preferences reflect the theoretical biases of each approach. The rule-based approach is rooted in deductive, rule-governed processes and has therefore focused on working-memory processes, which are traditionally associated with symbol manipulation. The experience-based approach, on the other hand, is interested in learning and has therefore focused on the associative mechanisms of LTM, traditionally associated with learning and the long-term storage of information.

However, this skewed approach to the human memory system is inconsistent both with experimental evidence on sentence comprehension and with what is now known about the psychology of memory. In Chapter 6, I describe Ericsson and Kintsch (1995) theory of LTM, which suggests that there is a close relationship between working-memory and LTM processes. The closeness of this relationship is inconsistent with the belief that working memory has a severely limited capacity – a central tenet of the rule-based approach. The relationship is also inconsistent with the idea that cognition is either rule-based or experience-based. In Ericsson and Kintsch's theory, rule-based and experience-based modes of cognition interweave in a single system. I will therefore use Ericsson and Kintsch's theory as a basis for integrating rule-based and experience-based approaches in psychological terms.

Psycholinguistics also makes use of mathematical constructs. For instance, generative psycholinguists use tree structure representations in modelling sentence comprehension processes. Connectionist approaches to psycholinguistics make use of networks in their models of sentence comprehension. Both trees and networks are actually abstract objects from a branch of mathematics known as graph theory. Although I will not be discussing graph theory in any depth in this book, I will show that it provides a mathematical formalism for handling both rule-based and probabilistic representations. I will therefore be proposing that graph theory can provide the mathematical underpinnings for an approach which integrates rules and experience.

When we consider foundational issues in psycholinguistics, therefore, we find good opportunities for integrating rule-based and experience-based perspectives, based on linguistics, psychology and mathematics. A further benefit of working with these fundamental conceptual issues is that we find ready-made theories of individual differences. Both Saussure's and Ericsson and Kintsch's theories provide accounts of individual differences. It might be purely a matter of coincidence that two quite independent theories which solve the problem of integrating rule-governed and experiential aspects of language also deal with the problem of individual differences. However, it could simply be that the duality of linguistic and other kinds of knowledge provides a natural fault line for dividing knowers into rule-followers and rote-learners; and theories that capture this duality appropriately are naturally primed to give a good account of individual differences.

I give brief descriptions of Ericsson and Kintsch's theory and Saussure's theory below. Extended descriptions are provided in later chapters. After showing how the two theories can be combined to provide a linguistically and psychologically plausible account of individual differences and other psycholinguistic issues, I give a brief description of graph theory and how it can be applied to the study of language and linguistic behaviour.

1.3 Ericsson and Kintsch's theory

Ericsson and Kintsch (1995) observe that psychologists have often assumed that human beings have a limited working-memory capacity. For instance, there is a severe limit to the number of digits or letters that experimental subjects can recall immediately after they have been read out to them. This inability has often been used as evidence that human beings have a fixed and limited working-memory capacity. It has also been assumed that these working-memory limitations constrain cognitive processes such as reasoning and the use of language. As we shall see in later parts of this book, proponents of the rule-based approach are more concerned with working-memory processes than they are with LTM processes. The reason for this interest is that cognitive processes are thought to manipulate information that is currently in working memory, where access is fast and reliable as opposed to information in LTM, where access is subject to interference and is therefore said to be slow and unreliable.

The hypothesised constraints of working-memory capacity on cognition form the psychological basis for the competence–performance distinction made in the rule-based approach. However, Ericsson and Kintsch point out that, while there may be grounds for positing a limited working-memory system, the total capacity of working memory is not determined by the amount of information, such as the number of digits or letters, so much as it is determined by the way in which that information is represented. For instance, Chi (1976) argues that children in early stages of literacy represent written words differently from literate adults. Children may represent words as sequences of discrete letters because their orthographic word representations are insufficiently developed. Adults, however, can represent words as single units. The fact that adults can recall more words than children is therefore not owing to their larger working-memory capacity but simply because they represent information in a more economical way.

Furthermore, there is evidence to show that when children and adults are tested on tasks in which the children are more skilled than the adults, the children display a greater working-memory capacity. For instance, if child chess experts are compared with adult chess novices in terms of their ability to recall information about chess, then the children recall more information. In general, it has been shown that the development of expertise in a particular domain is accompanied by increases in working memory for information in that domain. This superior working-memory capacity, however, is only specific to the domain of expertise, and experts and novices will display similar levels of working memory for information outside the domain.

The distinction between rule-based and experience-based knowledge which I have been discussing so far therefore appears to have a bearing on working-memory capacity. According to Ericsson and Kintsch, representing information in a systematic manner enables systematic storage of that information in LTM. In turn, systematic storage allows for systematic and therefore fast and reliable retrieval. It appears that the property of systematicity combats the interference which otherwise afflicts storage and retrieval from LTM. Systematicity therefore appears to confer a double advantage: it increases the abstract ability to represent a domain of reality and it also increases the working-memory capacity to process information from that domain.

For instance, according to Ericsson and Kintsch, skilled memorists can increase their STM for digits by up to 1000 per cent. They do this by creating stable systems of cues, which Ericsson and Kintsch refer to as retrieval structures. One type of retrieval structure is a hierarchical associative configuration of cues. The superordinate nodes in the hierarchy serve as cues for the subordinate nodes, which in turn serve as cues for their subordinates and so on, with the terminal nodes serving as cues for the actual digits. Digits are stored or retrieved after navigating the network in order to arrive at the appropriate memory locations.

The ability of retrieval structures to facilitate recall leads to an important observation, namely, that structured representations lead to what may be called rule-governed behaviour. Ericsson and Kintsch do not make this observation but it seems an obvious one to make. In navigating a retrieval structure for the purposes of storage or recall, the skilled memorist behaves in a manner which can be called rule-governed. In fact, this behaviour can be modelled in terms of the rewrite grammars that were used in early versions of generative grammar and that are still used in some psychological models of sentence comprehension.

A rewrite grammar consists of a set of symbols which successively replace each other to produce a sentence. For instance, an initial symbol, S (sentence), can be rewritten as the phrasal categories NP (noun phrase) and VP (verb phrase). These two symbols, in turn, can be rewritten as the terminal symbols or lexical categories Det (determiner), N (noun) and V (verb). Finally, the terminal symbols can be rewritten as actual lexical items, to produce a sentence, such as 'The cat slept.' In the same way, one can think of navigating a retrieval structure for recall purposes as a process of rewriting cues. Thus the terminal node is rewritten as its subordinate nodes, which, in turn, are rewritten as *their* subordinate nodes, and so on, until the terminal nodes are rewritten as the actual items of data to be recalled.

The fact that retrieval structures can be equated to rewrite systems is very important. It shows that what is called rule-governed behaviour can be seen as the *procedural* aspect of structured representations. It is therefore possible to reconcile rules with associative networks. Rules are simply descriptions of the process of navigating associative networks. Therefore, it could be that there is no necessary tension between rules and associations. Any apparent conflict might arise purely from viewing the same phenomenon as a process or as a structure.

It should be noted that retrieval structures serve the specific purpose of recalling digits or other information. In other domains of skill, it is necessary not only to be able to store and recall large amounts of information but also to manipulate that information in various ways. Ericsson and Kintsch refer to the associative structures created for this dual purpose as 'knowledge structures'. We can think of knowledge structures simultaneously as models of a domain of expertise as well as mnemonic aids for information from that domain. There are parallels here with computer software. Computer programs consist not only of rules for manipulating inputs but also special purpose data structures for storing information required for the computations. One could therefore think of knowledge structures metaphorically as a kind of mental software for carrying out specific mental tasks.

The use of highly efficient knowledge structures by experts can bring about dramatic gains in performance. Among the most spectacular demonstrations is the ability of some individuals to play and win up to ten games of blindfold chess simultaneously. This feat requires the chess expert to maintain the representations of ten chess boards in memory; update each board when a move is made and explore alternative moves in order to select winning moves. This ability is apparently not owing to innate talent in chess or to an inherently large working-memory capacity but to the development of highly efficient knowledge structures for chess.

Demonstrations such as these show that working-memory capacity does not have a fixed and severely limited capacity, as has been thought in the past, but that its total capacity depends on the systematicity of representations. This view effectively stands the competence–performance distinction on its head. It suggests that constraints on cognitive performance do not arise, in the final analysis, from limitations in working-memory capacity but from the efficiency of the system of representation used to encode information. It is therefore apparently the level of an individual's competence that constrains their performance and not the other way round. Proponents of the rule-based approach are therefore mistaken in assuming that native speakers possess a uniform grammatical competence whose expression is constrained by a limited working-memory capacity.

According to Ericsson and Kintsch, individual differences in sentence comprehension do not arise from inherent differences in working-memory capacity but from differences in the way in which linguistic information is represented. Poor comprehenders tend to represent linguistic information in small and relatively isolated units. This has negative consequences for both comprehension and recall, which require systematic relationships between units of information. Good comprehenders, on the other hand, create more integrated, global linguistic representations which facilitate both comprehension and recall. The representation and processing of language is therefore regarded by Ericsson and Kintsch as simply an instance of more general psychological mechanisms that also underlie skilled performance in diverse domains. As we shall see below, this global vision of cognitive organisation is mirrored and elaborated in Saussure's description of language.

1.4 Saussure's theory

According to Saussure, language is not innate to humanity, rather, what is innate is 'the faculty of constructing languages'. In his theory, described in Chapter 7, natural language is only one of many systems of representation created by humans. He called these systems of representation 'sign systems' and their purpose is to impose order on reality. The word 'model' conveys a sense of what sign systems are: they do not provide photographic renderings of reality, rather they represent selective aspects of reality for certain cognitive purposes. According to Saussure, all sign systems can be studied under one science which he called 'semiology' – the science of signs. Further, according to Saussure, sign systems, including natural language, have a common form of organisation.

A key element of this organisation is a distinction between what Saussure called *la langue* and *la parole*. Though Saussure applied this distinction specifically to the description of language, he intended it to have a broad application to all sign systems. He described *la langue* as the abstract system of language in the brain and *la parole* as the audible or written linguistic forms produced by a language community. However, *la langue* and *la parole* are not entirely distinct, as can be appreciated by considering in more detail how Saussure describes *la langue*. According to Saussure, *la langue* is an associative network consisting of linguistic units connected by two types of links – commonly referred to as paradigmatic and syntagmatic. Paradigmatic links connect units that share some abstract feature. For instance, all nouns share the property of being nouns and so they are linked paradigmatically. Syntagmatic links connect linguistic units into sequences, so that adjacent letters in a word or adjacent words in a phrase are linked syntagmatically.

While there might seem to be a clear distinction between *la langue* and *la parole*, it is important to recognise that a private copy of *la parole* is retained in memory. When linguistic units are linked syntagmatically, the links do not disappear after the act of speech but are stored permanently in LTM. According to Saussure, a syntagmatic link becomes stronger over time if two units co-occur frequently. So much so that, in Saussure's theory, frequently co-occurring units become unitised, forming larger units. These larger units can then enter into paradigmatic relationships with each other, giving rise to higher levels of the linguistic system.

Stored sequences can also serve as 'types' or templates for the production of novel sequences via analogy. *La langue* thus integrates its own product, *la parole*, within itself in spiralling levels of organisational complexity. *La langue* and *la parole* are therefore not completely distinct: each serves as input into the other in a process that makes the linguistic network progressively more elaborate.

The simple descriptive apparatus involving paradigmatic and syntagmatic links enables Saussure to handle, albeit only embryonically, both rule-governed and probabilistic aspects of language. The set of paradigmatic links encodes the system of abstract categories which comprise a mental grammar. Syntagmatic links, on the other hand, encode the sequential and probabilistic relationships between linguistic units. According to Saussure, all linguistic activity can be described in terms of paths across the associative network via the two kinds of links. For instance, the production of an utterance requires an individual to navigate the associative network paradigmatically, seeking linguistic units which satisfy certain search criteria. Selected units are then linked syntagmatically to generate a linguistic utterance which is then transmitted by the articulatory organs. The result of this process is a speech form, which belongs to the domain of *la parole*.

Earlier, I made a distinction in mathematical knowledge in terms of rule-based versus experience-based knowledge. This distinction also parallels Saussure's distinction between *la langue* and *la parole*. I suggested that the distinction between rule-based and experience-based knowledge of mathematics leads to individual differences in terms of rule-followers and rote-learners. Saussure himself did not directly discuss the possibility of individual differences, apart from suggesting that linguists might have more systematic knowledge of language than ordinary speakers. However, another great linguist, Roman Jakobson, applied Saussure's theory directly to individual differences along the lines of the paradigmatic–syntagmatic distinction.

According to Jakobson, individuals are, for a variety of reasons, biased towards the use of either paradigmatic or syntagmatic links. He suggested that some individuals prefer to carry out paradigmatic operations while others prefer syntagmatic operations. Without disagreeing with Jakobson, I would adapt his proposal as follows. The individuals who I refer to as rote-learners tend to rely greatly on *pre-established* syntagmatic links and to comprehend novel sentences by analogy to sentences stored in memory. Rule-followers, on the other hand, tend to make greater use of the grammatical categories encoded in paradigmatic connections. Their knowledge of language is therefore more abstract and categorical, providing them with a greater generative capacity. It is on account of their more abstract grammatical representations that rule-followers can understand complex sentences better than rote-learners.

As Chapter 8 will make clear, the dichotomy between rule-followers and rote-learners is not as sharp as it might seem from the above. The same individual can behave in one mode or the other with different aspects of language. Chapter 8 will provide a more precise formulation of the nature of individual differences. The key argument of the chapter is that individuals differ in the precision with which they define their linguistic units. As a result, some individuals develop a systematic knowledge of language while others end up with a more item-based knowledge. This pattern of individual differences is observed in first language development and in various aspects of adult performance. I propose that individuals differ in the ability to understand complex sentences because they define their syntactic units differently. In other words, individuals differ in their ability to understand complex sentences because they differ in syntactic competence and not because they differ in working-memory capacity.

In Chapter 9, I describe a series of experiments that I carried out to determine if variations in adult native speaker performance arise from variations in working-memory capacity (Just and Carpenter, 1992) or variations in syntactic representation (c.f. Ericsson and Kintsch, 1995). I tested three groups of 18-year-old native speakers of English on the comprehension and recall of English Complex NP constructions. Two groups were Low Academic Achievement (LAA) students and one group was High Academic Achievement (HAA) students (the reasons for the choice of subjects will be discussed in the chapter). The results showed that pre-test HAA levels of recall and comprehension were much higher than LAA levels.

One LAA group then underwent training to improve sentence recall and then carried out a post-test involving recall and comprehension. Levels of recall, but not comprehension, rose to HAA levels. Comprehension improved, however, when the LAA group was subsequently given comprehension training. The second LAA group underwent training to improve comprehension and then carried out a post-test involving recall and comprehension. Levels of both recall and comprehension rose to HAA levels. I conclude that, contrary to the working-memory hypothesis, the results support the view that variations in adult native speaker performance arise from variations in syntactic representation.

1.5 Implications

In Chapter 10, I examine the theoretical and educational implications of an account which integrates insights from the theories of Saussure and Ericsson and Kintsch and from graph theory in mathematics. More specifically, I consider the theoretical and educational implications of combining rules and experience for various sub-fields of linguistic research. I also consider the theoretical and educational implications of this integrated viewpoint with regard to the human intellectual potential. Finally, I consider the theoretical and educational implications of applying graph-theoretical techniques to the study of language. I describe these three sets of implications briefly below.

For linguistics, a key implication is that language needs to be described in both formal and probabilistic ways. A number of researchers are aware of this need (see, for instance, Bod, 1998; Wray, 2002). For the psychology of language, the implication is that models of sentence comprehension need to accommodate dual modes of comprehension and to allow for individual differences in bias towards one mode or the other. Again, this is not a new observation, and I review two hybrid models in Chapter 3. The view presented here therefore adds more weight to the need to integrate rule-based and probabilistic approaches in language-related research. The pattern of individual differences described in this book also has important methodological implications for both linguistics and experimental psycholinguistics. It indicates that the use of personal native speaker intuitions as primary data is highly unsafe. In addition, it also needs to be recognised that the almost exclusive use of university student subjects in psycholinguistic experiments results in a distorted view of the linguistic abilities of a population of native speakers.

In the case of language acquisition, it would seem that there is an important role for linguistic input, though there are also processes of rule-formation which cannot be handled in purely probabilistic terms. Yet again, this is not a novel observation. Bates *et al.* (1988) provide evidence which shows that children differ in terms of being analytical or holistic learners. The analytical learners tend to construct more systematic linguistic representations than holistic learners, who tend to possess piece-meal and less systematic representations. As Bates and colleagues point out, however, it is not as though children can simply be classified as analytical or holistic learners. Instead, the same child can display both tendencies in different aspects of language. Chapter 8 discusses this issue. The discussion in this book indicates that this pattern of individual differences is extremely general. It can be understood in terms of Saussure's paradigmatic and syntagmatic connections and it has implications for working-memory capacity in terms of Ericsson and Kintsch's theory.

The integrated view also has significant educational implications. Language teaching has tended to alternate between rule-based and usage-based pedagogical approaches. The discussion in this chapter has indicated that linguistic knowledge has both systematic and experiential aspects. There are certain irregular aspects of language that need to be learned by rote, such as idiomatic forms. At the same time, there are certain other aspects, such as phrase structure, that are best learned as rules, because they are highly regular. It is not therefore a question of deciding between a rule-based approach and an experience-based approach, but one of deciding which approach is appropriate for a given aspect of language. In Chapter 8, I present the argument that the efficient use of language depends on the precision with which linguistic units are defined. Fostering appropriate levels of precision requires learners to have their attention directed to certain aspects of language. I will present some ideas in Chapter 10 on how learners' attention can be directed towards syntactic features of language. I will also present ideas on how probabilistic aspects of language can be acquired in ways that avoid the tedium of language drills.

The second implication of combining insights from Saussure and Ericsson and Kintsch's theories has to do with the human intellectual potential. In order to appreciate this implication, it is necessary first to consider how the human mental potential is described by rule-based and experience-based approaches. On the one hand, the experience-based approach attributes human beings with a vast associative memory. However, the use of this resource is said to be severely constrained by the supposed inability of the human mind to create structured representations. On the other hand, the rule-based approach attributes human beings with an infinite capacity to create structured representations on the basis of rule-based knowledge. However, it is thought that the ability to create these representations is constrained by a severely limited working-memory capacity. In a way, therefore, each approach seems to take away with one hand what it gives with the other.

The integrated view combines the best of the two worlds and excludes the worst. Recall that, according to Saussure, human beings have an innate ability to construct languages. As we will see in Chapter 7, these languages have the capacity to create highly complex mental structures. According to Ericsson and Kintsch, on the other hand, LTM has a vast capacity. The combination of the two theories therefore suggests that human beings have an infinite abstract capacity to create structured representations as well as a vast associative memory in which to create those representations. It therefore appears that human beings have a vast computational potential which only requires efficient systems of representation in order to express itself.

The implications of this view are too enormous to encompass in a few words. It suggests that limits in human intellectual performance are largely artefacts of the systems of mental representation that are employed. The question arises about the extent to which psychological experiments tell us about inherent features of the human mind or about the systems of representation being used by experimental subjects. It could be that ascribing limits to the human intellectual potential is like ascribing the limitations of a computer program to the computer itself. One also wonders to what extent psychology might be better off experimenting with ways of increasing human cognitive performance rather than seeking to define its limits in terms of limits on working-memory capacity. From an educational perspective, the general implication seems to be that educational systems are content with levels of cognitive performance that are vastly below what students are potentially capable of. I will suggest, in Chapter 10, a general approach, in which language plays a key role, by which this potential can be unlocked.

1.6 Graph theory

Both Saussure and Ericsson and Kintsch indicate that language has the structure of an associative network. Yet again, this is not a unique view. The idea that language is represented mentally as a network is found in various forms in the work of Hockett (1985) and Lamb (1966, 1999). Currently, the notion of language as a network can be found in the work of Goldberg (1995), Halliday (1985), Hudson (2000) and Langacker (1987). Lamb's (1966) Stratificational Grammar and Hudson's (1984) Word Grammar, in particular, elaborate on the notion of language as a network.

As I mentioned earlier, the conceptual tools for describing networks are provided by graph theory, a field of mathematics which is concerned with the description of relationships (see Wilson and Watkins, 1990). Graphs consist of elements called nodes and links between elements known as edges. There can be any number of nodes in a graph and any number of links between nodes. There are some links called loops, which connect a node to itself. The nodes represent some elements of reality and the edges represent relationships between those elements. Edges can also be associated with a weight in order to indicate the importance of a relationship between the elements represented by the linked nodes. Weights can be used, for instance, to represent the strength of an association between two linguistic units. Graph theory provides a rich vocabulary for describing relationships in terms of graph properties. This vocabulary can be useful in both theoretical and educational contexts.

In theoretical contexts, the use of graph theory can enable linguistic analysis and psychological modelling to be carried using the same technical language. A recent impetus for using graph theory has been provided by Barabási (2002), who shows that networks are ubiquitous in nature and human institutions. Networks appear to have general properties that manifest in a wide variety of contexts, from the structure of molecules to the structure of the world-wide web. Barabási's approach inspired research by Motter *et al.* (2002), which showed that linguistic networks are like social networks in that they are densely interconnected. Whereas social networks are said to have six degrees of separation, linguistic networks appear to have, on average, three degrees of separation. For instance, the words 'universe' and 'actor' are linked semantically as follows: universe → nature → character → actor.

Motter *et al.*'s research was carried out by studying links between entries in a thesaurus and it is not clear how it relates to actual mental lexicons in individual human brains. Nevertheless, this kind of research is interesting because it links language with other fields of study. Studying language as a network can therefore make it possible to share results and techniques with more established sciences such as physics, biology and mathematics. This would be a healthy development in view of the current isolated state of linguistic research. In addition, graph theory would make it possible to use the same technical language in order to carry out both linguistic description and psychological modelling. Pioneering research along these lines has been carried out by Paul Meara and colleagues in modelling the mental lexicon (see Wilks, 1999).

The techniques of graph theory can also be applied to educational contexts. Firstly, graph theory provides a systematic way of describing language in terms of both its systematic and its probabilistic aspects. Graph theory can also make it possible to produce visually informative linguistic representations. Secondly, graph-theoretical techniques can be used as a form of linguistic measurement, in order to tap into the mental representations that students construct while learning a language. For instance, Wilks (1999) developed a graph-theoretical technique to measure the degree of connectivity in L1 and L2 lexical networks. There are many other graph properties in addition to connectivity that can potentially be used to tap into students' mental representations of language. In short, graph theory can provide a descriptive tool that has both theoretical and educational applications.

1.7 Summary

In summary, this book is concerned with native speaker variations in understanding complex sentences. It seeks to determine whether these variations are due to working-memory or syntactic competence. The answer that I provide to this question is that native speakers vary in competence. More specifically, it is possible to make a gross distinction between native speakers and to classify them, at least with respect to specific aspects of language, as rule-followers or rote-learners. This pattern of individual differences has implications for linguistic and psycholinguistic theories, which have been built on the assumption that knowledge of language is either rule-governed or experience-based. The difference between these two approaches cuts across many sub-fields of linguistic research and can be traced ultimately to rationalist and empiricist philosophical traditions. An account of individual differences which integrates these opposing perspectives therefore needs to take place at a foundational conceptual level. I propose that Saussure, Ericsson and Kintsch and a branch of mathematics called graph theory provide the concepts required for such an integrated account. I also explore some theoretical and educational implications of the integrated view.

2
Finite State and Generative Models

2.1 Introduction

This chapter introduces two basic descriptions of knowledge of language: the experience-based approach (for example, Elman, 1992; Hockett, 1955) and the rule-based approach (for example, Chomsky, 1957, 1965; Fodor and Pylyshyn, 1988). The experience-based approach describes knowledge of language as a network of associations between linguistic units. This network is derived from experience through associative learning and is stored in long-term memory (LTM). The rule-based approach describes knowledge of language as an innate formal system. The system is thought to guide the manipulation of linguistic symbols in working memory in order to derive syntactic structures. This chapter traces the conceptual influences surrounding the development of each approach, focusing on the earlier models of each type.

I will argue that Hockett's finite state model is based on associationism and information theory, while Chomsky's generative model is based on rationalism and a computer metaphor of mind. It is argued that these influences led Hockett to identify language with Saussure's *parole*, and Chomsky to identify it with Saussure's *langue*. It is further argued that Hockett and Chomsky's models represent a fragmentation of Saussure's more comprehensive conception of language. Two main consequences of this fragmentation are described and it is argued that these consequences also affect current experience-based and rule-based models. The following paragraphs describe Hockett's (1955) finite state model, which prefigures current connectionist experience-based models (for instance, Elman, 1992) and Chomsky's (1957) generative model, which prefigures current rule-based models (for instance, Gibson and Thomas, 1999). The description of each of the early models is preceded by a discussion of the intellectual influences surrounding its development.

2.2 The experience-based approach

2.2.1 Background to Hockett

Boeree (2000) describes associationism as the theory that the mind is composed of basic elements called ideas. These elements are organised by various laws of association. Aristotle proposed four such laws. The law of *contiguity* states that objects or events which co-occur in space or time become linked in the mind. The law of *frequency* states that the more frequent the co-occurrence the stronger the association. The law of *similarity* states that similar ideas will evoke each other. Finally, the law of *contrast* states that opposite ideas will also evoke each other. Subsequent associationists either sought to change the number of laws or suggested new ones. Some associationists focused on the laws of contiguity and frequency to the exclusion of everything else. Other associationists, however, added the notion of an active reason to the laws of association. Aristotle, for instance, proposed that reason guides strategic search and retrieval processes in order to recollect information from memory. Aristotle also proposed that reason actively structures and interprets incoming sensory input during the act of perception (see Anderson and Bower, 1980).

Hockett's era was dominated by two systems of ideas that were highly congruent with the brand of associationism which focused narrowly on the laws of contiguity and frequency. Hockett's mentor, Leonard Bloomfield, subscribed to logical positivism, an empiricist philosophy of science which held that 'all scientifically meaningful statements are translatable into physical terms' (Bloomfield, 1936: 325). This view translated into an exclusive focus on speech data to the exclusion of psychological factors. To some extent, this view was motivated by methodological considerations: 'In the division of labour, the linguist deals only with the speech signal [...] he is not competent to deal with problems of physiology or psychology.' (1933: 32). However, there were purely theoretical considerations as well. Bloomfield subscribed to behaviourism, a radical theory of psychology which denied the reality of mental constructs. In his view, the language user had no ideas but only words, which had a 'trigger effect upon the nervous system of his speech-fellows' (1936: 365). Consequently, he described knowledge of language as knowledge of the set of linguistic expressions generated by a speech community: 'The totality of utterances that can be made in a speech-community is the *language* of the speech-community' (Bloomfield, 1926: 27).

2.2.2 The finite state model

It is not clear how far Hockett subscribed to Bloomfield's views. However, the model of language which he developed was based exclusively on the laws of contiguity and frequency. The model is derived from Claude Shannon and Weaver's (1949) mathematical theory of information. Shannon described information not in terms of the raw number of messages transferred between sender and receiver but in terms of the predictability of a message. An important consequence of this description was that telecommunications systems could be made more efficient by exploiting statistical regularities in streams of messages. For instance, if a given sequence of messages tends to occur with a high degree of frequency, then that whole sequence can be recoded as a single message and encoded for transmission using a single electrical signal. This kind of coding is known as block coding (Pierce, 1980).

Shannon found that sequences in English text display marked statistical regularities which make it possible to create a statistical model of English text. Such a model could be derived through a quantitative analysis of a corpus of English text in order to specify the probability of encountering a given character or word given that certain other characters or words had already been encountered. Such a set of *transitional probabilities* can be represented in a compact form using a mathematical object called a finite state machine. Finite state machines, in turn, can be described in terms of a more general mathematical object called a graph. Graphs will reappear later in the book, so a brief description now is useful.

As described in Chapter 1, graph theory is a field of mathematics which deals with relationships between objects (Wilson and Watkins, 1990). Graphs consist of elements called nodes, and links between elements known as edges. The nodes represent some element of reality and the edges represent relationships between those elements. There can be any number of nodes in a graph and any number of links between nodes. There are some links called loops, which connect a node to itself. Edges can also be associated with a weight in order to quantify the strength of a relationship between the elements represented by the linked nodes.

Finite state machines can be represented using a graph. The nodes in the graph represent states of the machine and the edges represent the possibility that the machine can make a transition from one particular state to another. If state transitions are probabilistic, then weights on the edges are used to represent the probability of the machine transiting from one state to the other. A finite state machine can represent a statistical model of language in the following way. The states of the machine represent words and the weighted edges represent the probability of encountering certain other words when one particular word has already been encountered. This is essentially the sort of model which Hockett proposed. In Hockett's own words (1955: 7):

> GHQ [Grammatical Headquarters] can be in any of a very large number of different states. At any given moment, it is necessarily in one of these states. Associated with each state is an array of probabilities for the emission of the various morphemes of the language [. . .] When some morpheme is actually emitted, GHQ shifts to a new state.

The model could create new sequences by looping through certain states. For instance, the model could transit from the state representing the word 'old' back to that state. Such loops permit iterative operations (for example, the 'old old man') and therefore make it possible to generate a potentially infinite number of new sentences.

The fact that Hockett's model could generalise beyond observed data is important because the experience-based view is often characterised as one that treats knowledge of language as a mere list of previously encountered sentences (see, for instance, Fodor and Pylyshyn, 1988). In fact, Hockett sought to create a psychological model of the language user which could 'predict what other utterances the speaker of a language might produce' (1948: 279). Such predictions would have to be 'capable of passing the test of causal acceptance by the native speaker' (1952: 98). The model would therefore need to be able to 'generate any number of utterances in the language' (1954: 398) just as 'the speaker can produce any number of new utterances from essentially the same system of underlying habits' (1952: 98). Hockett's model was, in theory, capable of generalising to novel sentences, as suggested in the previous paragraphs. It did, however, contain serious limitations which were later exposed by Chomsky (1957). The next section will describe Chomsky's model before turning to these criticisms.

2.3 The rule-based approach

2.3.1 Background to Chomsky

Chomsky's model was explicitly based on rationalism (see Chomsky, 1965). Rationalism opposes associationism on a number of points.

Firstly, rationalists have always been deeply suspicious of experience and memory. Plato used his famous allegory of the case to argue that perceptions are unreliable and that they differ between individuals. He also argued that memory is only a remnant of the original experience and is therefore equally unreliable. He argued that knowledge of any real importance is innate; it is inaccessible to the sense organs and is accessible only to reason. He also argued that experience is only useful as a means of triggering dormant innate knowledge that has somehow been forgotten. Subsequent rationalists have suggested that what is innate are principles which constrain the manner in which sensory inputs are interpreted. The principles that have been suggested include the distinction between self and others, the principle of causality, the ability to reason in accordance with the laws of logic and the predisposition to project two-dimensional retinal images into three-dimensional mental representations (see Anderson and Bower, 1980).

Rationalism is closely associated with what is often referred to as a computer metaphor of mind. According to this metaphor, aspects of the mind/brain correspond to components of digital computers. For instance, human knowledge is construed as a set of programs that run on the brain in much the same way that computer software runs on computer hardware. An important idea arising from this conception is that these programs need to be co-ordinated by a central program called the central executive (Neisser, 1967). According to Anderson and Bower (1980), the central executive is sometimes informally associated with notions of free will, purpose and intention. The notion of a central executive is also sometimes associated with short-term memory (STM) processes. For instance, in their multi-stage model of human memory, Atkinson and Schiffrin propose that information is processed and maintained in STM via conscious control processes. Another example is Baddeley (1990), who identifies the central executive with working memory.

2.3.2 The generative model

Chomsky's (1957) model characterises natural language in terms of formal deductive systems. Such systems consist of a set of symbols and axioms which the rules manipulate in order to produce theorems. The rules can also operate on theorems to produce other theorems. Derivation is the name of the symbolic process by which theorems are produced from axioms or other theorems. In deductive systems such as geometry, axioms may be basic statements which define points, lines, and so on. In deductive systems called formal languages, axioms and theorems can simply be symbols or sequences of symbols. It is important to point out that there is nothing essentially rationalist about the use of deductive systems. Bloomfield, for instance, cast his theory of language in the form of an axiomatic system. Harris, a student of Bloomfield, took a step further and proposed that a corpus could be modelled using an axiomatic system. He suggested that the process of corpus systematisation can be made to yield a set of statements which

> ...form a deductive system with axiomatically defined initial statements and with theorems concerning the relations among them. The final theorems would indicate the structure of the utterances of the language in terms of the preceding parts of the system (1951: 372).

Chomsky extended this idea and proposed that the linguistic knowledge possessed by the language user can itself be characterised as a formal language. Chomsky's (1957) model consists of a set of strings of symbols and a set of rules which manipulate the symbol strings to ultimately produce terminal strings (sentences). The symbol strings are syntactic symbols such as S, NP, VP, Det, N, V and so on. Phrase structure rules rewrite the initial strings into other strings. For instance, S can be rewritten as NP VP; NP can be rewritten as Det N, which in turn can be rewritten as 'the child'. Similarly, VP can be rewritten as V, which can be rewritten as 'slept'. The string 'the child slept' is called a terminal string because it cannot be rewritten into any other string.

The derivational history of this string has a hierarchical structure referred to as the phrase structure of the sentence. An important feature of the phrase structure rules proposed by Chomsky is that they are recursive, so that a symbol which occurs earlier in the derivational history can recur later in the derivation. For instance, S is rewritten as NP VP. NP can also be rewritten as N S (in cases where a noun is postmodified by a clause). The rules are recursive because S dominates NP but it is also dominated by NP. The process of recursion can continue indefinitely, allowing the generation of arbitrarily complex phrase structures.

There is also another set of rules called 'transformational rules'. These rules are used to transform strings derived by phrase structure rules into other strings. For instance, the string of symbols – NP_1 Aux V NP_2 – can be written into the terminal string 'The boy fed the dog'. This terminal string is said to be in the active voice. It is possible to transform this string into the passive voice by applying transformational rules to the string of syntactic symbols which precede the terminal string. Thus NP_1 Aux V

$NP_2 \rightarrow NP_2$ be+en Aux V NP_1. This second string can be rewritten as 'The dog was fed by the boy' which is a passive voice paraphrase of the earlier active voice sentence.

Hockett and Chomsky's models therefore differ considerably in their characterisation of knowledge of language. It is important to note, however, that Hockett and Chomsky's views represent a fragmentation of the ideas of Ferdinand Saussure, who predates both. The brief description of Saussure's theory below helps clarify the nature of the difference between Hockett and Chomsky's models. A more detailed description is provided in Chapter 7.

2.4 Descriptive limitations of each approach

2.4.1 A brief preview of Saussure's theory of language

Saussure saw language as only one of the many forms of representation constructed by the human mind. An important feature of these forms of representation, in his view, is that they are self-enclosed systems which define their constituent units via sets of contrasts. The basic units of Saussure's representational systems are signs. A sign is an association between a concept or 'signified' and a phonic image or 'signifier'. Signs are related to each other in two basic ways. Paradigmatic relationships pertain to relations of similarity between signs. For instance, lexical classes form paradigmatic sets. Syntagmatic relationships pertain to structural relations between signs which comprise a more complex sign. Thus, morphemes in a word and words in a phrase comprise syntagmatic sequences.

Saussure's theory therefore includes the associative laws of contrast, contiguity and similarity. Chapter 7 will show that Saussure's theory also incorporates the law of frequency, which is responsible for the unitisation of frequently occurring sequences and for the induction of categories from linguistic experience. It will be argued in that chapter that the laws of contiguity and frequency give Saussure's theory the potential to describe the kinds of phenomena captured by experience-based models, while the law of similarity gives the theory the potential to describe the kinds of phenomena captured by rule-based models.

Saussure regarded paradigmatic and syntagmatic relationships as complementary facets of language. Paradigmatic sets offer representational choices to the language user while syntagmatic relationships enable the user to combine simple signs into more complex signs. Syntagmatic sequences, in turn, also comprise paradigmatic sets containing units of greater complexity. For instance, a word is a syntagmatic sequence of morpheme selections, yet words form paradigmatic series in their own right, in the form of lexical classes. This means that a sequence of choices can be stored as a ready-made group of signs which then forms a single choice. The storage of syntagmatic sequences in memory facilitates the process of constructing representations by cutting out the need to actively construct representations already resident in memory. This idea parallels the use of block coding in information theory in order to increase the efficiency of a telecommunications system. It also parallels the use of formulae in mathematics.

The relationship between the paradigmatic and syntagmatic dimensions underlies Saussure's distinction between *langue* and *parole*. For Saussure, *langue* is the abstract and individual expressive potential while *parole* is the tangible and public set of expressions generated by *langue*. It appears that Hockett focused on *parole* whereas Chomsky focused on *langue*. However, in Saussure's theory, *langue* and *parole* interact. The manner of this interaction can be observed by noting the following correspondences. Firstly, *langue* corresponds to the paradigmatic dimension in that both offer expressive choices. *Parole*, on the other hand, corresponds to the syntagmatic dimension in that both pertain to combinations of specific choices. There is also a close relationship between *parole* and both the syntagmatic and paradigmatic dimensions of *langue*. Firstly, each individual language user stores a subset of *parole* in LTM in the form of the syntagmatic sequences which that individual has heard from others or produced by themselves. At the same time, *langue* is latent in *parole*. This observation is based on the fact that, according to Saussure, *langue* is learnt by attending to speech. Saussure also states that *langue* changes in adulthood in response to patterns of *parole*. Therefore, *langue* and *parole* interact as do the paradigmatic and syntagmatic dimensions of *langue*.

By breaking this interaction, experience-based and rule-based approaches render themselves unable to explain certain aspects of linguistic knowledge adequately. Experience-based models tend to describe language in purely statistical terms. This is true of Hockett's (1955) model and Elman's (1992) model, which are described in Chapter 4. Rule-based models, on the other hand, tend to describe language in purely formal terms, as in Chomsky's (1957) model and Gibson and Thomas' (1999) model. The following paragraphs argue that these forms of description are inadequate. Firstly, the finite state model 'undergenerates' in the sense that it does not generate all the possible sentence forms of English. Secondly, the generative model 'overgenerates' in the sense that it generates sentence forms which native speakers of

English do not recognise as sentences of English. The generative model also overgenerates in another sense. The model defines syntactic rules over lexical categories. The assumption is that all native speakers can assign each word to its appropriate lexical category. However, this appears to be an overestimation of native speaker lexical knowledge. Thirdly, the generative model also 'undergenerates' because it cannot generate idiomatic expressions.

2.4.2 Undergeneration in the finite state model

This section argues that, by focusing exclusively on the laws of contiguity and frequency, Hockett's finite state model renders itself incapable of supporting complex structured representations. A demonstration to this effect was provided by Chomsky (1957) and it will be shown to apply with equal force to current connectionist models. Chomsky's criticism was based on an example of a simple language with just two words, *a* and *b*. The grammar of that hypothetical language is such that any sequence of words is always followed by a mirror image of that sequence, for instance, *aa, ab, abba, aaaa, bbbb, aabbaa, abbbba*, and so on. Such a grammar is not defined in terms of contiguous relationships between words but in terms of the overall structure of word sequences, that is, the structural requirement is that the second half of a sequence must be a mirror image of the first half. It is not clear how sequential learning can capture this regularity (however, see an attempt to prove otherwise by Wiles and Elman, 1995).

Chomsky then argued that there are syntactic structures in English which have this mirror image property. The most important of such mirror image structures is the classical self-embedded sentence. In this type of sentence, the subject noun of the matrix clause is post-modified by a subordinate clause whose subject noun is also post-modified by a subordinate clause and so on, for instance, *The man who the doctor who the nurse called treated died*. This sort of sentence has a mirror image property in the sense that each subordinating clause nests within it a subordinate clause so that the sequence of the predicate verb phrases is the reverse of the order of the corresponding subject nouns. Chomsky argued that the existence of self-embedded sentences in English shows that syntactic structures in English are recursive and therefore open to an arbitrary degree of structural elaboration. Given that Hockett's finite state machine could not represent mirror image structures, Chomsky concluded that such a machine could not serve exclusively as a model of English grammar or of users of English. The same argument has been used by Fodor and Pylyshyn (1988) against connectionist models. However, this argument turns out to be a double-edged sword because it poses serious problems for Chomsky's (1957) model as well.

2.4.3 Overgeneration in the generative model

Reich (1969) observes that, in developing his argument for the existence of recursive rules in English, Chomsky (1957) stated that English sentences can have the form:

(a) If S_1, then S_2

Supposing that S_1 is realised as

(b) If your knight takes my pawn, then your bishop would take my knight.

Then, it is possible, using recursive rules, to realise (a) as

(c) If if your knight takes my pawn, then your bishop would take my knight, then I should move my king.

Applying recursive rules to a greater extent, Chomsky's argument also licenses the following realisation of (a).

(d) If if if your knight takes my pawn, then your bishop would take my knight, then my queen would take your pawn, then you should move your king.

(c) and (d) are clearly consistent with the general form (a) and, by definition, must be considered grammatical sentences of English. Yet, Reich argues, such sentences are unacceptable to native speakers. The problem that Reich identifies here is that of 'overgeneration': generative grammars create sentences which, while technically grammatical, are difficult to describe as natural language sentences (see Christidis, 1985). The problem of overgeneration suggests that a set of rules cannot, by itself, provide a description of what native speakers know of their language. The problem shows that native speakers know what sorts of sentences occur in their language and what sorts of sentences do not, even if both sorts of sentences conform to the putative rules of the grammar. Presumably, native speakers can discriminate between normal and overgenerated sentences on the basis of their memory of those sentences which they have heard before. In that case, it needs to be acknowledged that knowledge of language may consist of both rules and previous linguistic experience.

Christidis (1985) notes that proponents of generative grammar sometimes seek to banish the products of overgeneration by using devices called 'filters' or 'surface structure constraints'. For instance, Perlmutter (1971) proposed a '*that-that*' filter in order to exclude sentences such as:

> *That that she smokes bothers you surprises me* (that is, 'It surprises me that it bothers you that she smokes.').

While the use of filters may be justified on descriptive grounds, the use of filters in psychological models would imply that language users go through the process of producing or understanding overgenerated sentences, but are prevented from completing the process at the very last moment by the operation of a filter. In fact, current rule-based psychological models do not incorporate filters and simply pass over the problem of overgeneration in silence. There also appear to be two sorts of problems associated with the processing of overgenerated sentences. One problem is that of comprehending such sentences. Once the structure is explained, however, comprehension is possible. For instance, many native speaker subjects are unable to understand self-embedded sentences until they are taught to do so (see Chapter 3). A second problem is that, even when one can understand an overgenerated sentence, it still doesn't 'sound right'. This second problem could simply be a question of frequency: if used often, such structures might end up sounding normal.

A related problem is that an individual with incomplete or inaccurate information about a lexical item will be hampered in two sorts of ways. Firstly, if that individual does not assign the correct lexical category to a lexical item, they may make errors in assigning the correct syntactic structure to a sentence. After all, syntactic rules are defined over lexical categories. Secondly, imprecise knowledge of the syntactic environment associated with a given lexical item can also be expected to lead to possible errors in the assignment of syntactic structure. For instance, an individual whose knowledge of the verb 'fear' is confined to those contexts in which the verb takes a noun complement (for example, 'I fear the dog') may find it difficult to understand a sentence in which the verb has a sentential complement (for example, 'I fear the dog will be starving by now'). There is, in fact, some evidence that college students who differ in lexical knowledge also differ in syntactic processing, such that individuals with poor lexical knowledge appear unable to assign phrase structure to sentences (see Cupples and Holmes, 1987, 1992). The issue here is that describing language in purely formal terms does not provide sufficient information about what specific individuals actually know about their native language.

2.4.4 Undergeneration in the generative model

It turns out that the generative model also undergenerates. As noted by Gross (1979), no natural language has ever been completely described in terms of rules. Instead, generative models of language often involve a distinction between the core and the periphery of the grammar. The core relates to those aspects of a language which yield to formalisation. The periphery relates to idiomatic expressions which resist formalisation. These expressions are simply listed in the lexicon because they cannot be generated by rule. Schachter (1996) estimates that three-quarters of any pedagogical grammar belong to the periphery, though she does not say how she arrives at that estimate. In any case, a brief perusal of descriptive grammars such as Quirk *et al.* (1972) will show that there is a mass of grammatical facts which cannot be derived from a simple set of principles. The generative model therefore undergenerates, in the sense that it cannot generate all the grammatical sentences of a language.

2.4.5 The dual nature of linguistic knowledge

Chomsky's argument concerning mirror languages indicated that Hockett's model also suffers from the problem of undergeneration because they cannot handle self-embedded sentences. Fodor and Pylyshyn (1988) have also raised the same argument against connectionist models, which are also finite state. The response of proponents of finite state models to this criticism is that human beings cannot process self-embedded sentences beyond a certain degree of embedding. For instance, Reich claimed that multiply self-embedded sentences do not occur in natural language and therefore the inability of finite state machines to handle them was not a flaw. A similar argument is made by proponents of current connectionist models (for example, Christiansen and Chater, 1999; Elman, 1992), who argue that self-embedding beyond three levels is not found in natural languages.

However, this reply is inaccurate. Sampson (2000) carried out a corpus study which shows that multiple self-embedding beyond three levels can occur in natural language. Experimental evidence will also be reviewed in Chapter 3 which shows that subjects can be trained to understand and produce self-embedded sentences whose complexity far exceeds the limits displayed by connectionist models. While it may be true that high levels of self-embedding are not a very common feature of natural language, the fact that subjects can be trained to use recursive *rules* means that a model of sentence processing which cannot handle such rules is seriously limited. It should also be noted that the fact that subjects have to be *trained* to use such rules is also problematic for the rule-based approach, given that this approach regards the use of recursive rules as an automatic aspect of normal native speaker competence.

It therefore appears that finite state models underestimate the human potential for language. By itself, the human ability to engage in mathematical thought indicates that Chomsky (1957) and Fodor and Pylyshyn (1988) are correct to insist that human beings have the *potential* to create representations of arbitrary complexity. However, describing this potential in purely formal terms leads proponents of rule-based models overestimate what human beings actually know about language. But while rule-based models have a strong tendency to overgenerate, they undergenerate because they cannot generate idiomatic expressions, which, despite comprising a considerable proportion of language, are consigned to the periphery of descriptive efforts. The limitations of each approach can be illustrated by drawing analogies between linguistic and mathematical knowledge. Mathematical knowledge is partly formal and partly experiential. Individuals differ to the degree that they understand the formal aspects of mathematics and to the degree that they have rote-learned mathematical formulas. There is evidence that individuals differ in linguistic knowledge along the same lines. Such differences are difficult to explain from either a purely rule-based or experience-based approach. In fact, as argued below, both Hockett and Chomsky appeared to assume that native speakers possess a uniform grammatical competence in their native language.

2.4.6 The assumption of uniform syntactic competence

Hockett and Chomsky's belief in uniform native speaker competence is shown by their faith in native speaker intuition. Both proposed that their models would be validated if the generated sentences were 'capable of passing the test of casual acceptance by the native speaker' (Hockett, 1952: 98) or 'acceptable to a native speaker' (Chomsky, 1957: 13). The assumption that native speakers do indeed possess a uniform competence is, however, much more closely associated with the rule-based than with the experience-based approach.

While language users might conceivably possess *different* generative grammars, the overwhelming assumption has been that if language users possess a generative grammar, then it must be the same generative grammar. This assumption appears to be based on the additional assumption that the generative grammars supposedly possessed by native speakers are similar to those proposed by linguists. Here again, the chain of reasoning is weak because linguists have proposed numerous grammars and there are no clear criteria to decide which among them is the most psychologically plausible. Another possible line of reasoning is that knowledge of language is innate, therefore all speakers of the same language possess the same grammar. But the line of reasoning here is also weak because genetically inherited characteristics generally display considerable natural variation. In short, although the assumption of uniform competence is widespread and deeply entrenched, its logical underpinnings are vague. Chapters 3 and 8 will show that the assumption has no empirical underpinnings either.

It does appear, however, that the assumption is related to proposals by Chomsky. For instance, Chomsky (1966: 75) states:

> By a 'generative grammar', I mean a description of the tacit competence of the speaker-hearer that underlies his actual performance in production and perception (understanding) of speech. A generative grammar, ideally, specifies a pairing of phonetic and semantic representations over an infinite range; it thus constitutes a hypothesis as to how the speaker-hearer interprets utterances, abstracting away from the many factors that interweave with tacit competence to determine actual performance.

Elsewhere, however, Chomsky (1965: 9) states:

> To avoid what has been a continuing misunderstanding, it is perhaps worth while to reiterate that a generative grammar is not a model for a speaker or a hearer. It attempts to characterize in the most neutral possible terms the knowledge of the language that provides a basis for actual use of language by a speaker-hearer.

Yet, in the same publication, Chomsky (1965: 59) states:

> ...it seems reasonable to suppose that a child cannot help constructing a particular sort of transformational grammar to account for the data presented to him, any more than he can control his perception of solid objects or his attention to line and angle.

It is difficult to decide from these quotations precisely what Chomsky's position was regarding the psychological status of generative grammar.

Many psychologists, however, have always taken the interpretation implied by the first quotation, which is that language users possess a syntactic competence which is obscured from view by performance factors such as working-memory capacity (see Gleitman and Gleitman, 1970 for an expression of this argument). Psychologists have also assumed that the level of competence is uniform among native speakers of a language. This psychological position has remained essentially unchanged from the 1960s to the present. This position is so deeply entrenched that most experiments in sentence processing are carried out on university-student subjects. This is a highly unrepresentative sample which has undergone a highly selective educational process. The only possible justification for using such an unrepresentative sample is an underlying assumption that all normal native speakers have the same underlying syntactic competence.

There are other reasons, however, apart from those inspired by Chomsky, for thinking that native speakers possess a uniform syntactic competence. Eysenck and Keane (1995) note that cognitive psychology marginalises individual differences in an effort to describe *general* information processes. A related consideration is that cognitive psychology treats cognitive processes, such as language processing, in terms of algorithmic symbol manipulation processes which can be modelled by formal systems. For some reason, this view leads to the assumption that these formal systems are the same across individuals. However, there are infinitely many possible algorithms for carrying out any given task. Consider, for instance, the multitude of computer languages and programs which carry out the same computational tasks. The fact that a cognitive process can be modelled by a formal system is no reason to think that such systems are the same for all individuals. The basis for marginalising individual differences in cognitive psychology is therefore as vague as that in linguistics and psycholinguistics.

The next section considers how each approach can deal with findings that individuals differ in syntactic knowledge.

2.4.7 The problem of individual differences

Neither Hockett nor Chomsky addressed individual differences in linguistic knowledge. If they had been faced with the problem of explaining such differences, then the assumptions underlying their models might have suggested the following sorts of explanations.

Hockett's model is experience-based, so it would be logical to suppose that individual differences in experience will lead to individual differences in the ability to process sentences. Chomsky's model, on the other hand, is strongly committed to a formal and innate knowledge of language. To suggest that individuals possess different formal systems would complicate matters considerably. For one thing, it would raise questions about the role of native speaker intuitions in linguistic theory. It seems more likely that Chomsky would have preferred to locate individual differences in linguistically extrinsic factors, such as individual differences in working-memory capacity. As it turns out, these explanatory routes, have, in fact, been taken by current experience-based and rule-based models. Thus MacDonald and Christiansen (2002) propose that variations in linguistic experience are a major source of individual differences. On the other hand, Just and Carpenter (1992) and Caplan and Waters (1999) both propose that individual differences are due to individual differences in working-memory capacity.

It is worth noting, however, that the rule-based approach does, in theory, allow for certain kinds of individual differences. For instance, the rule-based approach needs to acknowledge the possibility of individual differences in the *ability* to use rules. In Chapter 1, I made a distinction between rule-governed and experiential knowledge of highly regular domains such as mathematics. I will not repeat this argument here, except to point out, that even when individuals possess knowledge of the same rules, they will not necessarily use them in the same way. There is a story that a mathematics teacher needed to keep his students occupied for some time and so he devised a task which he thought would do the trick. He asked his students to calculate the sum of integers from 1 to 100. However, 10-year-old Friedrich Gauss found a quick way of doing the task. If the sequence is reversed (100, 99, . . . , 1), the sum of the corresponding entries is equal to the size of the sequence $(n)+1$:

$$\begin{array}{rrrrr} & 100 & 99 & \dots & 1 \\ + & 1 & 2 & \dots & 100 \\ = & 101 & 101 & \dots & 101 \end{array}$$

If 101 is multiplied by 100, that gives twice the sum of the original sequence. Therefore, all that is necessary to calculate the sum of the original sequence is to divide that sum by a half. Hence the formula $\frac{1}{2}n(n+1)$.

The point of the illustration is this. Doubtless, Gauss, his teacher and his fellow pupils all understood the rules of arithmetic. However, they clearly differed in their ability to use those rules. Gauss could see that the problem could be represented in a form which eliminated the need to carry out repetitive, time-consuming and error-prone steps. Therefore, despite having the same basic rules as the others, Gauss was able to represent the problem in a more efficient manner. It would be implausible to suggest that this ability is based on mathematically extrinsic performance factors, such as working memory. I will not attempt to define this ability, except to note that it involves, in part, the ability to perceive an equivalence between two expressions: $1+2+\ldots+n$ and $\frac{1}{2}n(n+1)$. The relevance of this illustration to the discussion is that, as a matter of principle, the fact that two individuals possess knowledge of the same set of rules does not entail the same ability to use those rules, and this difference in ability cannot be trivialised to an effect of performance factors.

Secondly, as noted in Chapter 1, it is not even the case that knowledge of language can be described exclusively in terms of rules. Rule-based models undergenerate. This means that there are many grammatical phenomena in a language which cannot be derived via rules. Yet such phenomena are an intrinsic component of native speaker competence. Such phenomena must presumably be learned in a piecemeal fashion, given appropriate linguistic experience. Given that individuals differ in linguistic experience, it is reasonable to suppose that they will differ in their knowledge of idiomatic phenomena. A related issue is that individuals also differ in lexical knowledge. This knowledge is also an intrinsic aspect of native speaker competence because it underlies the ability to apply syntactic rules, given that such rules are defined over lexical categories. As argued earlier, an inability to assign a word to its appropriate lexical category may lead to a failure in syntactic processing. Therefore, even if native speakers possessed the same syntactic rules, they would still differ in syntactic comprehension on account of individual differences in lexical knowledge.

The next section considers the ways in which the focus on either formal or statistical aspects of language by rule-based and experience-based models leads to an asymmetrical approach to the human memory system.

2.4.8 The asymmetrical treatment of human memory

Experience-based and rule-based models make different assumptions about the memory system primarily responsible for sentence processing. Hockett's model appears to require LTM rather than STM. As described earlier, the model represents the grammar of language as a learned set of transitional probabilities. Sentences are therefore represented as trajectories through state space. In addition, the model characterises linguistic knowledge purely in terms of the laws of contiguity and frequency. There is no provision for a central supervisor to control the process of representation. Therefore, the model does not actively construct representations: it simply traverses the associative network in a manner determined by the stored transitional probabilities and the current input. All that the model requires therefore is an associative network in LTM which encodes transitional probabilities between words. To some extent, this description is also true of connectionist networks (see Chapter 4). In general, the experience-based approach is more concerned with LTM processes than STM processes. This means that models developed within this approach tend to provide poor descriptions of STM processes, in particular, the control processes associated with STM.

Chomsky's model, on the other hand, appears to require STM. Long-term memory is clearly required to store the grammar and the lexicon. However, the act of syntactic derivation involves the construction of intermediate representations which require temporary storage. A short-term or working memory seems appropriate for this purpose. Proponents of the generative model from the 1960s onwards have always assumed that phrase structure representations are constructed and stored temporarily in STM. However, the almost exclusive preoccupation with STM processes means that models developed within the rule-based approach are not very good at describing LTM processes, such as learning and the use of previous linguistic experience during sentence comprehension.

Chapter 6 will review findings related to memory that were generated by a debate between proponents of rule-based and experience-based approaches to language comprehension. The aim of the debate was to decide whether the process of sentence comprehension creates LTM traces of the surface forms of sentences or not. If surface forms were not stored in LTM, then this would have suggested that sentence processing was essentially a STM process. It would also suggest that the experience-based approach was wrong because that approach requires LTM records of sentence processing episodes in order for learning to take place. On the other hand, if sentence processing episodes did create LTM traces of surface forms, then that would support the idea that knowledge of language includes knowledge of previously encountered sentences. It would also raise questions about the characterisation of sentence processing as a purely STM process. Therefore, what was potentially at stake in this debate was a key argument of either approach.

2.5 Parallels between early and current theories

The preceding section has described finite state and generative models of language. Along the way, various parallels have been noted between earlier and current models of each type. These parallels will now be summarised briefly for convenience.

The finite state model arises from an associationist view of psychology. It regards language as equivalent to the corpus of expressions generated by a speech community. This corpus can be modelled statistically using a finite state machine which can be implemented psychologically as an associative network in LTM. Finite state machines treat sentence representations as unique trajectories through state space. The finite state approach underestimates the linguistic potential of language users by failing to support recursive rules. Proponents of the approach justify this lack by arguing that recursive rules are not part of natural language (Reich, 1969).

This description applies almost word-for-word to current connectionist models (see description in Elman, 1992). While such models do differ in significant respects from Hockett's model, the fundamentals are the same. Connectionist models are associationist in character. They represent a statistical model of a corpus, referred to as 'the training set'. Connectionist models also treat sentence representations as trajectories through state space (Elman, 1992). Elman's model does not support recursive rules, though it can process a limited degree of self-embedding. As did proponents of earlier finite state models, proponents of connectionist models argue that recursive rules are not part of natural language (Christiansen and Chater, 1999).

The rule-based model arises from a rationalist view of psychology. It treats knowledge of language as a deductive system, which takes the form of a generative grammar. It supports recursive rules but overestimates the linguistic abilities of language users by assuming that the ability to use formal rules is automatic. The model treats sentence representations as symbol structures which describe the derivational history of the sentence in the form of phrase structure trees. The model requires STM in which to construct temporary, intermediate representations. It makes a fundamental distinction between competence and performance.

These observations apply equally to current rule-based models. These models adopt a rationalist view of psychology. They treat knowledge of language as a deductive system in the form of if-then productions (see Just and Carpenter, 1992). They support recursive rules (see Fodor and

Pylyshyn, 1988). However, the proponents of these models do not discuss the descriptive problems of simple generative grammars, such as the fact that such grammars both overgenerate and undergenerate. The models assume phrase structure representations of sentences. They also assume that native speakers possess a uniform competence which is only indirectly reflected in performance.

The earlier and current theories of each type are therefore identical in terms of certain key assumptions. These assumptions lead current experience-based models to attribute individual variations in comprehension primarily to individual differences in linguistic experience. This explanation arises naturally from the fact that knowledge of language is considered to be inherently statistical. The assumptions underlying the rule-based models also lead them to attribute individual variations to individual differences in the performance factor of working memory. This explanation is almost inevitable, given a strong commitment to a uniform competence grammar with recursive rules.

3
Early Experimental Studies

3.1 Introduction

This chapter reviews experimental studies into knowledge of language that were carried out to decide between Hockett's (1955) and Chomsky's (1957) models. Many of these experiments were carried out in the 1950s and 1960s. Subsequent research has focused largely on mechanisms of sentence processing, rather than on the knowledge underlying such mechanisms (see Mitchell, 1994). The experiments reviewed here therefore represent the most sustained attempt to find out what form knowledge of language takes in the individual human brain. The experiments show that language users are sensitive to the statistical regularities of language. Evidence for knowledge of phrase structure rules is inconclusive. However, there were experiments which did show that individuals can be taught to use recursive rules. I will argue that this last finding is deeply problematic for both experience-based and rule-based models.

The review consists of four sections, dealing with experimental evidence for (a) sensitivity to the statistical structure of language; (b) knowledge of phrase structure rules; (c) knowledge of recursive rules and (d) the existence of a uniform syntactic competence.

3.2 Evidence for statistical structure

The fact that language has a statistical structure has been known since ancient times. The ninth century Arabian code breaker, Al-Kindi, exploited the statistical structure of language to decode messages that had been encrypted using a method called the substitution code. The code works by systematically substituting one letter for another in order to render the message unintelligible. The intended recipient of the message, who is in possession of the code, can then decode the message by reversing the substitution process. Al-Kindi broke the code when he found that each letter in a written language occurs with a characteristic frequency. For instance, to use the example of English, 'e' comprises about 13 per cent of English text, while 'z' comprises less than 1 per cent. Al-Kindi compiled a table of letter frequencies from the written language and another table from the message text and discovered that he could work out the system of letter substitutions and thereby decode the hidden message. His technique was used to great effect by a 17th century code breaker to decode secret communications between Mary Queen of Scots and her co-conspirators, leading to the conviction and execution of Mary for plotting against Elizabeth I (see Singh, 2000).

3.2.1 Transitional probabilities

It was mentioned in Chapter 2 that Shannon and Weaver (1949) analysed the statistical structure of English text as part of the development of information theory. They studied not only the absolute frequencies of letters in text, as did Al-Kindi, but also the probability of one letter or word occurring immediately after another. This study showed that linguistic sequences display statistical regularities called transitional probabilities. An experiment by Hockett (1953) showed that human subjects are sensitive to transitional probabilities in language and that such probabilities can be used to determine the boundaries of linguistic units. In the rather informal experiment, the experimenter would suggest the first letter of the first word of a sentence. The participants would then have to guess the subsequent letters, right up to the end of the sentence. At each guess, the experimenter would say whether or not the guess was correct. On analysing the number of correct and incorrect guesses, it was found that the highest number of correct guesses occurred *within* words while the highest number of incorrect guesses occurred *between* words. Elman (1990) obtained a similar finding using a connectionist model. What this finding shows is simply that word boundaries occur at points of highest statistical unpredictability and that, therefore, there is a statistical basis for the segmentation of language into units (Hockett, 1953).

Formal experiments by Goldman-Eisler (1958) indicated that pauses in speech occur at points of high uncertainty. Maclay and Osgood (1959) also found that pauses occur more frequently at phrase boundaries than within phrases. In addition, Maclay and Osgood found that pauses occur more frequently before lexical words than before function words. Function words are generally more frequent than lexical words; therefore, this finding is additional evidence showing that language users are sensitive to statistical regularities in natural language text. Reading research undertaken at a later period indicates that skilled readers are more sensitive to transitional probabilities in orthography than less skilled individuals (for example, Lefton *et al.*, 1973; Scheerer-Neumann *et al.*, 1978). Muise *et al.* (1972) and Le-Blanc *et al.* (1971) carried out cross linguistic studies, which indicated that the speed of letter reading is related to transitional probabilities.

Related observations were made in the psychology of memory and perception. Much of this research was carried out using a measure called 'order of approximation'. An illustration of this concept is provided by Miller and Selfridge (1951). They constructed word sequences by giving subjects a certain amount of linguistic context and asking them to use that context in a sentence. For instance, a subject might be given the word 'he' and then be asked to compose a sentence beginning with that word. The first word used by the subject immediately following 'he' would then be given to another subject who would also be asked to continue the sentence, starting with that word. In this way, a set of two-word sequences was built up. Such sequences are called 'second order approximations' (first order approximations are obtained by compiling a random list of words). Third orders of approximation were obtained by giving subjects two words of context, fourth orders by giving them three words of context and so on, up to seven orders of approximation (see Table 3.1).

In a number of studies, it was found that the span of immediate memory (short-term recall) is a function of the order of approximation

Table 3.1 Seven orders of approximation

Zero-order approximation	Byway consequence handsomely financier bent flux cavalry swiftness weather-beaten extent
First-order approximation	Abilities with that beside I for waltz you the sewing
Second-order approximation	Was he went to the newspaper in deep and
Third-order approximation	Tall and thin boy is a biped is the beat
Fourth-order approximation	Saw the football game will end at midnight on January
Fifth-order approximation	They saw the play Saturday and sat down beside him
Seventh-order approximation	Recognise her abilities in music after she scolded him before
English text	The history of California is largely that of a railroad

to English. Memory span increased with increasing order of approximation, but reached asymptotic levels at the fourth order of approximation. It was also found that word lists in the seventh order of approximation were recalled just as well as text (Deese and Kaufman, 1957; Marks and Jack, 1952; Miller and Selfridge, 1951; Richardson and Voss, 1960; Sharp, 1958). A related finding was obtained by Miller (1958) and Aborn and Rubenstein (1952), who both observed that letter strings generated by a finite state grammar are learnt and recalled better than random strings.

There was also evidence for the influence of statistical structure on language perception. Traul and Black (1965) found that the intelligibility of speech increases and the variability of errors decreases with increasing order of approximation. Pollack and Pickett (1964) found that speech perception accuracy increases with increasing context. Miller *et al.* (1954) found that letter recognition accuracy is greater for letters in familiar sequences compared to letters in unfamiliar sequences. Finally, Onishi (1962) found that eye-voice span increases with increasing order of approximation while Imae and Takeuchi (1959) found that intelligibility, attention span, recall and eye-voice span all increase with order of approximation. It should be noted, however, that in some cases, order of approximation could have been confounded with non-statistical sources of linguistic organisation.

The research described above suggests that (a) language has a statistical structure; (b) language users are sensitive to this structure and (c) language users differ in sensitivity to statistical regularities in language. The finite state model therefore has a degree of psychological plausibility, notwithstanding the real limitations pointed by Chomsky (1957). The findings also suggest that language users store records of their linguistic experience in LTM. These LTM records appear to facilitate cognitive processes such as speech perception, reading, attention and short-term recall. These findings have an important bearing on the debate in that they also show that so-called performance limitations in cognitive processing may actual ly be indicators of insufficient learning. If so, then there is a logical possibility that individual differences in sentence comprehension are not due to individual differences in working-memory capacity, as is widely believed, but rather to individual differences in LTM for language. This issue is discussed more fully in Chapter 6.

The next section reviews evidence for the hypothesis that language users possess phrase structure rules which they use during the process of understanding sentences.

3.3 Evidence for phrase structure rules

3.3.1 Phrasal organisation

Chomsky's theory of generative grammar inspired a sustained attempt to prove the 'psychological reality of phrase structure rules' in the 1960s. The search for evidence for phrase structure rules consisted largely of attempts to determine if subjects imposed a phrase structure organisation on sentences. A large number of experiments were carried out to observe effects of syntax on (a) the intelligibility of sentences heard under noise; (b) sentence recall under various conditions and (c) click location.

Miller (1962), and Miller and Isard (1963) found that the presence of syntactic structure in a string of words increased the intelligibility of words under noisy conditions. Epstein (1961a,b) also found that syntactic structure facilitated the recall of pseudo-word strings. In a related study, Marks and Miller (1964) found that recall of grammatical sentences was superior to recall of ungrammatical sentences. Johnson (1965) devised a metric called 'transitional error probability' which showed that sentences tended to be recalled in intact phrase units. This result was taken to show that language users impose phrasal organisation on sentences. The result was partially replicated by Levelt (1970) using a different metric. However, Levelt also found that the phrase units in subjects' recalls did not always coincide with the phrase structures prescribed by linguistic theory.

Anglin and Miller (1968) found that passages which were segmented in conformity with phrase structure were memorised more rapidly than passages whose segmentation violated phrase structure. Graf and Torrey (1966) reported that comprehension was better when rapidly presented text segments conformed to major syntactic boundaries than when they conformed to minor syntactic boundaries. A related finding was made by Mehler *et al.* (1967), who discovered that eye-fixation patterns conformed to syntactic structure.

In click location studies, an auditory click was superimposed on a recording of a sentence at a phrase boundary or within a phrase. Subjects listened to the test sentence and then wrote it down, indicating where the click had occurred. Fodor and Bever (1965) observed a tendency for clicks to 'migrate' towards major syntactic boundaries even if the clicks had actually occurred well within a phrase. In addition, the greatest number of correct responses was obtained when clicks were located at major phrase boundaries rather than within phrases. These findings were replicated by Bever *et al.* (1969), Garret *et al.* (1966) and Holmes and Forster (1970).

Thus, there appears to be psychological evidence for phrase structure organisation. However, there was a fundamental problem with the experimental logic underlying all the studies which reported evidence for phrasal organisation. It had been hoped that evidence of phrase structure would provide support for the reality of phrase structure *rules*. However, this was a mistaken approach, because evidence of phrase structure did not, by itself, constitute evidence for the psychological reality of phrase structure *rules* any more than it provided evidence for other sources of phrasal organisation. Such organisation could arise, for instance, from the fact that transitional probabilities are stronger within phrases than between them (for example, Goldman-Eisler, 1958; Maclay and Osgood, 1959; see also Rosenberg, 1968).

There were also great difficulties in unconfounding syntax from other variables. In particular, it was difficult to separate syntax from semantics. It was quite possible that grammatical sentences were more intelligible and easier to remember than ungrammatical sentences simply because the grammatical sentences made sense whereas the ungrammatical sentences did not. For instance, Mandler and Mandler (1964) and Townsend and Saltz (1972) found that sentence recall tended to reflect propositional rather than syntactic structure. The use of pseudo-words by Epstein (1961b) could not solve this problem either, because (a) grammatical morphemes can enable a language user to assign a thematic structure to a string of pseudo-words (Levelt, 1974) and (b) subjects could make use of linear order constraints to interpret pseudo-word sentences without actually using phrase structure rules.

There were problems with the other experimental techniques. Loosen (1972, cited in Levelt, 1974) discovered that what appeared to be effects of syntax in sentence recall could be due to a response bias, whereby subjects imposed phrase structure during the production of responses rather than during comprehension. Other researchers found that apparent effects of syntax on click location were task effects (see Levelt, 1974) and that the same pattern of responses could be obtained using non-linguistic materials (Reber and Anderson, 1970). However, see Townsend and Bever (2001) for a recent response to these criticisms.

The fact that Bock and Loebell (1990) were still seeking a way to prove the existence of phrase structure rules is a good indication of how intractable the problem has proven to be. It is noteworthy that Fodor *et al.* (1974), who had led the attempt to prove the psychological reality of phrase structure rules, acknowledged that evidence for such rules had not been obtained, although there was evidence for phrase structural organisation. Bever (1970) went as far as to propose that sentence comprehension was based on statistically based sentence patterns which he called canonical sentoids (see also Bever, 1988; Townsend and Bever, 2001).

Miller (1962: 754) had observed correctly that an effect of phrase structural organisation on sentences 'does not show that some form of grammatical structure must be preferred to, say, a Markovian [finite state] structure of the sort that communication theorists talk about'. This observation is repeated by Miller and Isard (1963: 224) who found an apparent effect of syntactic organisation on sentence processing but noted that, 'It is not possible to discredit [the] Markovian model in terms of our present data.' The only way to decide between generative and finite state models, Miller and Isard concluded, was to study the processing of self-embedded sentences. As shown in Chapter 2, such sentences are not supported by Hockett's model. However, Chomsky's (1957) generative grammar does support the processing of self-embedded sentences via the application of recursive rules. Therefore, if it was observed that subjects could process self-embedded sentences, that observation would constitute evidence for the generative model and evidence against the finite state model. Experiments on the processing of self-embedded sentences are reviewed next.

3.3.2 Evidence for recursive rules

Miller carried out two experiments on the processing of self-embedded sentences. In the first experiment, Miller (1962) asked his subjects to repeat self-embedded sentences. He found that (a) subjects repeated the sentences with list intonation and (b) subjects could only recall about seven words from each sentence. These results were negative for the generative model in two ways. Firstly, the fact that subjects repeated the sentences with list intonation suggested that they were not imposing syntactic organisation on the sentences but were treating them as mere lists of words. Secondly, Miller (1956) had found that subjects could only recall around seven unrelated items in tests of short-term recall. The fact that subjects in the (1962) experiment could only recall around seven items therefore suggested, once again, that these subjects were treating the self-embedded sentences as mere word lists. However, Miller did not draw any conclusion from these results, despite the potential significance of the results in helping to decide between generative and finite state models.

In the second experiment, Miller and Isard (1964) presented subjects with six sentences which ranged from zero to four degrees of self-embedding, for example,

(0) She liked the man that visited the jeweller that made the ring that won the prize that was given at the fair.
(1) The man that she liked visited the jeweller that made the ring that won the prize that was given at the fair.
(2) The jeweller that the man that she liked visited made the ring that won the prize that was given at the fair.
(3) The ring that the jeweller that the man that she liked visited made won the prize that was given at the fair.
(4) The prize that the ring that the jeweller that the man that she liked visited made won was given at the fair.
(random) Won given that that the fair man made visited prize the at the she that jeweller was the ring that.

A recording of each sentence was presented five times and subjects attempted to repeat each sentence after each presentation. When the subjects' responses were analysed, it was found that performance improved over the five presentations, so that subjects recalled a sentence better after the fifth presentation compared to the first. Secondly, the sentences were recalled better than the random word strings. Thirdly, the greater the number of self-embeddings in a sentence, the less well it was recalled. A fourth result was that there was not much difference in recall between sentences with zero or one self-embedding. However, there was a marked difference in recall between sentences with one and two self-embeddings. There was also a marked difference in recall between sentences with two and three self-embeddings, but there was not much difference in recall between sentences with three and four self-embeddings.

Miller and Isard proposed an explanation for the fact that sentences with two or more self-embeddings were recalled less well than sentences with zero or one self-embedding. The explanation was based on an analogy to the way in which recursive procedures are executed by computers. Often, a computer program will 'call' another program, called a subroutine, to carry out a given task. When this happens, the operation of the main program is suspended while the subroutine is executed. The computer must therefore store the information (the return address) of the point at which execution of the main program was suspended. When the subroutine has been executed, the computer looks up the return address and execution of the main program continues. It sometimes happens that the subroutine will also call another subroutine to carry out part of the task. Sometimes, a subroutine will call itself to carry out a task. Miller and Isard proposed that when a subroutine in the human language processing system calls itself, the return address is erased.

In effect, what Miller and Isard proposed is that the human language processing system cannot process recursive structures. This proposal was endorsed by Chomsky (1965). It is also important to note that Miller and Isard (1964) did not attribute the difficulty of centre-embedded sentences to *memory capacity* as such, but rather, to something more akin to interference. The fact that Miller and Isard (1964) and Chomsky (1965) argue, in effect, that humans cannot process recursive rules appears to undermine Chomsky's (1957) argument against the finite state model. In that argument, Chomsky argued that the grammars of natural language contain recursive rules. He then argued that the finite state model does not support recursive rule and therefore it is incapable of modelling the linguistic abilities of language users. In his (1965) book, however, he proposes that the human language processing system *cannot* process recursive structures, not because it lacks sufficient STM, but because it is constitutionally incapable of handling recursion.

One could argue that the finite state model is simpler and therefore preferable to the highly speculative and *post-hoc* suggestion offered by Miller and Isard. Unaccountably, Miller and Isard did not interpret their results in favour of the finite state model. Recall that the motivation to study the processing of self-embedded sentences, as suggested by Miller (1962) and Miller and Isard (1963), was to decide between the generative and finite state models. However, no mention is made of the finite state model by Miller and Isard (1964). Instead, Miller and Isard (1964: 294) propose 'a distinction between knowing a rule and obeying it'. It is difficult, however, to understand how language users can be said to know recursive rules if the language processing system is inherently incapable of processing such rules. One possibility is that there is a complete disjunction between knowledge of language and the part of the brain that actually processes language.

It seems more sensible to suggest that the disjunction is between linguists' descriptions of language and what individual language users actually know about language. This point is consistent with Chomsky's purely methodological justification for recursive rules. In Chomsky (1956: 109), he states that

> the assumption that language are infinite is made for the purpose of simplifying the description. If a grammar has no recursive steps [...] it will be prohibitively complex, it will, in fact, turn out to be little

> better than a list of strings or of morpheme class sequences in the case of natural languages. It is does have recursive devices, it will produce infinitely many sentences.

In this passage, the use of recursive rules is justified on purely descriptive grounds. If that is the sole justification for the use of such rules, then Chomsky's (1957) argument against the finite state model appears less compelling. Linguists might find it economical to use recursive rules, but language users might well be content with finite state grammars, however complex and formally inelegant such grammars might be.

The real state of affairs appears to be more complex, however, and inexplicable purely in terms of either generative or finite state models. Experimental work subsequent to Miller and Isard (1964) and reviewed below, produced evidence that difficulties in comprehending self-embedded sentences were due to insufficient syntactic knowledge. However, language users could be taught to use recursive rules.

The first indications that problems with processing self-embedded sentences might be due to limitations in syntactic knowledge as opposed to limitations in the language processing system were provided by Blumenthal (1966). He asked subjects to paraphrase sentences with three degrees of self-embedding. Examples of a test sentence and incorrect responses by subjects are shown below:

Test item	The manager whom the designer whom the typist whom the receptionist encourages interests consults phoned the producer.
Error 1	The manager that the designer encourages, that the typist interests, that the receptionist consults, phoned the producer.
Error 2	The manager that the designer, the typist, and the receptionist encourage, interest, and consult, phoned the producer.

Only 41 out of 160 responses were correct. In the remaining cases, subjects 'either [(a)] interpreted the sentences as having one relative clause containing a compound subject and a compound verb, or [(b)] they perceived three successive relative clauses, all referring to the initial noun' (453). Blumenthal concluded that such errors were due to ignorance of the fact that self-embedded sentences are based on a valid grammatical form in English. Marks (1968), for instance, found that sentences with more than one degree of self-embedding were considered ungrammatical by native speakers of English. It should be noted, however, that the studies by Blumenthal and Marks did not disprove that processing limitations might be responsible for comprehension difficulties. The studies simply indicated that alternative explanations were possible.

A stronger argument that the problems in processing self-embedded sentences had a linguistic source was made by Stolz (1967). Stolz proposed that a distinction be made between weak and strong productive capacity in characterising an individual's grammatical knowledge. Weak productive capacity is of the sort which allows an individual to comprehend only those structures for which he or she has learnt a specific rule. On the other hand,

> A person would be considered productive in a stronger sense if he needed to know only the relative clause construction and then was able to apply it recursively on demand by using it along with a non-specific sort of second-order metalinguistic principle of recursiveness applicable to any substantive linguistic rule having certain properties. By displaying this sort of productivity, a speaker would give evidence that he possessed these abstract grammatical constructs in a readily manipulable and relatively autonomous form not much different than their counterparts in linguistic theory (1967: 869).

Stolz further proposed that, under conditions where time and memory pressures were minimal, an individual could either (a) comprehend self-embedded sentences immediately, thus displaying strong productivity; or (b) comprehend them after encountering a sufficient number of examples with non-linguistic cues to their interpretation, thus displaying a weak productivity or (c) process the structures incorrectly, thus displaying a lack of productivity.

A straightforward interpretation of the generative model would lead one to expect that subjects would behave in terms of (a). (It should be noted, however, that this prediction is inconsistent with the more complicated proposals of Miller and Isard (1964) and Chomsky (1965), which suggest that language users are *incapable* of understanding sentences with more than one self-embedding.) The finite state model could not support the processing of self-embedded sentences, unless the test sentences were exact replicas of sentences stored in the model. Therefore, the finite state model would predict that subjects would behave in terms of (c). Possibility (b) is predicted by neither the finite state model nor the generative model, either in its simpler version

(Chomsky, 1957) or in its more complicated version (Chomsky, 1965; Miller and Isard, 1964).

Stolz asked subjects to rewrite each self-embedded sentence as a series of simple sentences with appropriately matched subjects and verbs. The sentences were presented with or without semantic support and feedback. For instance, the first sentence below is semantically supported while the second is not:

> *The stone that the boy that the club members initiated threw hit the window.*
> *The dog that the cat that the bird fought scolded approached the colt.*

Feedback was provided by telling the subject the correct interpretation to each sentence after the subject had attempted to interpret the sentence and before the subject moved on to the next sentence. Subjects were divided into four groups. One group received both feedback and semantic support; another received semantic support but no feedback. A third group received feedback but no semantic support and a fourth group received neither feedback nor semantic support. There were two blocks of sentences. The first block contained the conditions described above. In the second block, there was neither semantic support nor feedback.

The result for the first block was that performance was lowest when neither semantic support nor feedback was given. There were no significant differences between the other three conditions. In the second block, performance was highest for those subjects who had been given feedback without semantic support in the first block. There were no significant differences between the other three conditions. This result is interesting because it shows that subjects displayed a high degree of comprehension only when given training in the form of feedback. Stolz (1967: 872) concluded:

> This study would indicate that many decoders can display productivity of only the weaker sort. That is, these listeners can decode a novel recursive sentence only when its recursive quality has been learned as a specific part of the specific grammatical rule involved. They would not, then, show evidence that recursiveness has the status of a readily applicable, autonomous concept.

This result is consistent with possibility (b) above, which, it was argued, is predicted by neither the generative nor the finite state model. With respect to the effects of semantic support, Stolz concluded:

> When a sentence involving restrictive semantic constraints is encountered, the average S[subject] seemed to do very little syntactic processing. It may be generally the case that relatively unfamiliar syntax is decoded only insofar as an analysis is absolutely required to produce a semantic interpretation of a sentence. If there is only a single reasonable interpretation of a sentence due to semantic (or pragmatic) constraints, then detailed syntactic processing may be largely by-passed (872).

Stolz's findings and conclusions bear significantly on current views about the relationships between syntactic processing, semantic processing and working-memory capacity. Just and Carpenter (1992) propose that individuals with low working-memory capacity cannot use semantic constraints during sentence comprehension. Caplan and Waters suggest along similar lines that individuals with low working-memory capacity have difficulty with processing the propositional contents of sentences. Current views about the role of semantics in sentence comprehension are therefore inconsistent with Stolz's finding that many subjects could not comprehend self-embedded sentences purely on the basis of syntactic structure. Chapter 4 will also describe an experiment by Gibson and Thomas (1999) which, the authors argue, demonstrates an effect of working-memory capacity on syntactic processing. It will be suggested that the experiment confounds semantic constraints with syntactic structure. Stolz's study is a good illustration of the consequences for current research of failing to take earlier experimental work into account.

Stolz's findings were partially replicated by Freedle and Craun (1970) and Powell and Peters (1973). Freedle and Craun used a prompt sentence to help subjects understand self-embedded sentences, for example:

Prompt *The woman that the boy complimented despised* someone.
Test The policeman that *the woman that the boy complimented despised* arrested the drunk.

They devised two lists of materials. List A consisted of a prompt plus test pair (as above) and test sentences with one, two or three self-embeddings. List B consisted solely of test sentences with three self-embeddings with no prompt sentence provided. Group A was presented with List A first and List B second; Group B was presented with List B first and List A second. Subjects were asked to paraphrase the test sentences. The results for List A for both groups were as follows. Group

A achieved 100 per cent paraphrase accuracy for test sentences with one and two self-embeddings and 80 per cent paraphrase accuracy on sentences with three self-embeddings. Group B achieved 100 per cent accuracy on sentences with one degree of self-embedding; 76 per cent paraphrase accuracy on sentences with two self-embeddings and 33 per cent paraphrase accuracy on sentences with three self-embeddings.

The results indicate that subjects had no difficulty understanding sentences with one self-embedding but had some difficulty with sentences containing two or more embeddings. The results also indicate that subjects were able to *induce* the structure of self-embedded sentences from the prompt sentence. The results from Group B indicated that prior exposure to multiply embedded sentences without a clue to their structure interfered with subjects' ability to make use of such a clue when it was later provided. A possible reason for this is that Group B subjects had found a method of interpreting multiply self-embedded sentences (incorrectly) and continued to use this method even if evidence from List A suggested that the method was incorrect.

Freedle and Craun also found evidence of individual differences in the ability to generalise from two embeddings to three in Group A:

> [. . .] The fact that not all Group A Ss responded appropriately when triple-SE [triply self-embedded] sentences were given (even though they correctly re-wrote the aid sentences) suggests that these Ss had not yet discovered the general rule governing SE structure but instead, had discovered a "rule" specific to the number of embeddings in the aid sentences (249).

Chomsky (1957) had considered the possibility of constructing a grammar which specifies the number of self-embeddings. He had rejected this possibility because it would have resulted in a complex grammar. It seemed simpler and more elegant to use recursive rules which applied generally without reference to the degree of self-embedding. It is therefore interesting that some subjects in Freedle and Craun's experiment induced a recursive rule which specified the number of self-embeddings. This finding highlights the disjunction between linguistic theory and the representation of language in the individual human mind. Linguists may generalise maximally for descriptive purposes but language users are not under the same pressure. Freedle and Craun's experiment indicates that language users differ among themselves in the generality of the rules that they induce from linguistic experience.

The studies by Stolz and Freedle and Craun showed that native speakers do not necessarily possess recursive rules but that they can learn them. This result turns out to be problematic for both generative and finite state models. The generative model attributes to native speakers an automatic ability to use recursive rules. The fact that subjects have to learn to use such rules is therefore problematic. It was noted earlier that, in view of the difficulties displayed by language users in comprehending self-embedded sentences, Miller and Isard (1964) and Chomsky (1965) had proposed a highly speculative explanation which implied that the human language system is constitutionally incapable of handling recursion. The experimental data, rather perversely, showed that subjects *can* learn to use such rules. Both the pro- and anti-recursion versions of the generative model were therefore at odds with the experimental data.

The data is also problematic for the finite state model. That model could only process recursive sentences that it had stored. The model cannot therefore explain the ability of subjects to generalise immediately from a single embedding to two or three self-embeddings (in the Freedle and Craun experiment). In addition, subjects were generalising to new sentences. It is not clear, of course, whether the subjects in Stolz and Freedle and Craun's experiments did indeed learn a recursive rule which they could apply to an arbitrary level of structural complexity. It is possible that these subjects could have induced a recursive rule which they could only apply to a maximum of three levels of self-embedding.

However, Blaubergs and Braine (1974) found that subjects could be trained to understand sentences with five degrees of self-embedding and to produce sentences with up to seven levels of self-embedding. This level of complexity is far beyond that attained by connectionist networks (see Elman, 1992; Christiansen and Chater, 1999). The results also suggest that subjects had, in fact, learned a recursive rule which could be reapplied to arbitrary levels of syntactic complexity. In line with this possibility, Blaubergs and Braine found that trained subjects performed differently depending on the mode of presentation. When a visual display was used, subjects could understand sentences with up to five degrees of self-embedding. However, when the sentences were presented auditorily, subjects could only understand sentences with two degrees of self-embedding. Presumably, the drop in performance was due to subjects having to carry out both processing and storage functions simultaneously. This result shows that there is at least one clear instance in which it is justified to make a distinction between competence and performance.

To conclude this section, it may be noted that, as stated earlier, Chomsky's (1957) argument concerning self-embedded sentences turned out to be a double-edged sword. It presents problems for both rule-based and experience-based approaches. The rule-based approach attributes to language users a strong productive capacity. However, untrained subjects display weak productive capacity. Their performance limitations are therefore a direct reflection of their syntactic competence. Experience-based approaches attribute to language users a weak productive capacity. However, language users can be trained to acquire strong productive capacity. This result is problematic because the experience-based models do not support recursive rules.

It therefore appears that each approach describes only part of a more complex picture. An important element of this more complex picture is that individuals vary in the sorts of generalisations which they derive about language. The next section briefly reviews experimental investigations into the idea that native speakers possess a uniform syntactic competence. A more detailed review is provided in Chapter 7.

3.3.3 Evidence for uniform competence

The section on recursive rules has already reported individual differences in knowledge of recursive rules as well as in the ability to learn such rules. This evidence was produced by Blumenthal (1966), Freedle and Craun (1970), Powell and Peters (1973) and Stolz (1967). As noted above, even Miller and Isard (1964) reported evidence of individual variations. However, there is also evidence of individual differences in grammatical knowledge at much lower levels of structural complexity. Differences have been observed at phonological, morphological and clause levels of language.

Ferguson (1979) reports an informal experiment which uncovered wide variations in knowledge of phonological rules among native English speakers. These variations were found in both adults and children. He also reports more formal studies which indicate that some children tend to build up their phonological representations in a more systematic manner than other children. Maratsos (1976) and Biber (1983) separately found individual differences in knowledge of the definite and indefinite article systems in English and Somali speakers, respectively. Schnitzer (1993) found, rather surprisingly, that children displayed greater control of verbal inflectional morphology in Spanish than adults. Gleitman and Gleitman (1970) and Geer *et al.* (1971) found systematic individual differences between groups of English speakers in the ability to understand three- and two-word noun compounds.

Cupples and Holmes (1987) found individual differences in word-class knowledge between groups of native English speakers. This is a significant finding which needs to be studied further. The significance lies in the fact that, if individuals differ in lexical knowledge, then they may also be expected to differ in syntactic processing, given that syntactic rules are defined over lexical categories. Cupples and Holmes do, in fact, report findings which suggest that an inability to assign lexical categories accurately results in an inability to impose phrasal organisation on sentences. There are independent studies, described in the following paragraphs, which show that native English speakers differ in their ability to impose phrasal organisation. It is difficult to say, with some of these findings, if this difficulty is related to lexical knowledge. There are other studies, however, where there is a possible link between lexical knowledge and syntactic processing.

Cromer (1970) found individual differences in the effects of textual layout on the ability to impose phrasal organisation (but see Oakhill, 1994). Levin and Kaplan (1970) found that the eye-voice span of skilled readers extended to phrase boundaries more often than those of less skilled readers. Finally, Muncer and Bever (1984) found that errors made by skilled readers in letter detection were more sensitive to phrasal boundaries than those of less skilled readers.

A study by Carol Chomsky (1969) appears to link lexical knowledge more clearly with syntactic processing. Chomsky found differences among five- to ten-year-olds in the ability to comprehend sentences like,

(a) I told Bill where to go.
(b) I asked Bill where to go.

Sentence (a) obeys what Chomsky referred to as the minimal distance principle in that the subject of the non-finite clause ('to go') is the noun phrase which most immediately precedes the verb ('Bill'). Sentence (b) does not obey the minimal distance principle, since the subject of the non-finite clause is the noun phrase which is most distant from the verb. Chomsky found that some children interpreted the two types of sentences in the same way: they thought that in (b) the subject of the non-finite clause is 'Bill'. There is a lexical difference between (a) and (b) which is centred on the main verb. With verbs like 'told', the object of the main verb is the subject in the infinitival clause. With verbs like 'asked', the subject of the main verb is also the subject of the infinitival clause. However, the reverse may also be true, where 'asked' involves making a request or issuing a command, as in 'I asked Bill to leave'. It is possible that some of the children may not have been aware of the distinction between the two types of verbs and assumed that 'asked' always behaves like 'told'.

Chomsky's findings were criticised by Kessel (1970) on methodological grounds. However, the results were replicated by Sanders (1971), who found considerable individual differences among adult native English speakers in the ability to comprehend exceptional structures. She concluded that:

> At least one structure thought to be commonly known to adults was not consistently understood by almost half the subjects in the study. It is not unreasonable to speculate that there may well be other syntactic structures which have been assumed to be part of ordinary adult language, but which most adults may not understand (744).

Kramer *et al.* (1972) carried out a longitudinal study involving exceptional structures. They traced a number of individuals from childhood to college level and found that certain individuals never learned to comprehend the structures correctly. They concluded that: 'It appears that for every two competent adults, there is one noncompetent adult, even in [the] college group' (125).

Apparently on the strength of Kramer *et al.*'s study, Noam Chomsky (1980: 175–176) acknowledged that individuals vary in syntactic competence and not just in performance:

> [...] I would be inclined to think, even without any investigation, that there would be a correlation between linguistic performance and intelligence; people who are intelligent use language much better than other people most of the time. They may even know more about language; thus when we speak about a fixed steady state, which is of course idealized, it may well be (and there is in fact some evidence) that the steady state attained is rather different among people of different educational level [...] it is entirely conceivable that some complex structures just aren't developed by a large number of people, perhaps because the degree of stimulation in their external environment isn't sufficient for them to develop.

The idea that native speakers have a uniform syntactic competence is therefore unfounded on empirical grounds. Chapter 2 also showed that the idea has exceedingly tenuous theoretical underpinnings. The implication of this conclusion is that current rule-based theories which attribute individual differences exclusively to individual differences in working-memory capacity need to revise their position. Individuals do appear to differ in grammatical competence, whatever differences may or may not exist in working-memory capacity.

3.4 Summary

In summary, the experimental attempts to decide between generative and finite state models showed that (a) language has a statistical structure and language users are sensitive to it; (b) language users vary in their sensitivity to the statistical structure of language; (c) knowledge of the statistical structure of language helps language users reduce limitations in performance usually attributed to intrinsic limitations of the cognitive machinery; (d) it is not clear whether or not language users possess phrase structure rules; (e) language users can employ recursive rules but they differ in their ability to do so; (f) language users can be trained to use recursive rules and (g) language users differ in syntactic competence. These results indicate that neither the generative nor the finite state model provides a complete picture of the form that knowledge of language takes in the minds of language users. The next chapter will argue that this conclusion applies equally to current rule-based and experience-based models of sentence comprehension.

4
Connectionist and Symbolic Models

4.1 Introduction

This chapter describes two current experience-based and rule-based models of sentence comprehension (Elman, 1992; Gibson and Thomas, 1999). The description covers the background of each theory, which takes the form of connectionism in the case of Elman (1992) and symbolic processing, in the case of Gibson and Thomas (1999). The key point made in this chapter is that current experience-based and rule-based models make the same assumptions that Hockett and Chomsky's models made about the nature of linguistic knowledge and its relationship to the human memory system. It will be argued that the current models are therefore subject to the same limitations as the earlier models. It will also be shown that current models do not address the experimental findings reported in Chapter 3. The discussion in this chapter lays the groundwork for the discussion of models of individual differences in Chapter 5. These models are either directly based on the models described in this chapter or closely related to them.

This chapter is divided into two sections. Section 4.2 describes connectionism in general and Elman's model in particular. It is shown how this model makes the same assumptions as that of Hockett (1955). Section 4.3 describes symbolic processing with reference to the theories of Fodor (1983) and Fodor and Pylyshyn (1988). The section then describes the model of Gibson and Thomas (1999).

It is useful to begin by stating that connectionism and symbolic processing are two opposing sets of very general ideas about how the mind works. Connectionism is closely related to associationism while symbolic processing is closely related to rationalism. The manner in which proponents of either approach describe linguistic knowledge is therefore part of a much wider vision of the nature of the mind and the nature and acquisition of knowledge. While this section deals with connectionism and the next deals with symbolic processing, it will be necessary to make references to symbolic processing in this section and to make references to connectionism in the next section. This is because the two approaches take systematically opposing views on key issues and these views are best understood in relation to each other.

4.2 Connectionism

Ellis and Humphreys (1999) provide a summary of some key characteristics of connectionism. The following paragraphs describe five of these characteristics. Four characteristics will be illustrated later in the discussion with reference to specific connectionist models and related to the connectionist account of individual differences in the comprehension of complex sentences described in Chapter 5.

Firstly, connectionism explores the ability of brain-like mechanisms to carry out computations. Some connectionists attempt to model real neurons directly while others create more abstract models whose relationship to the brain is less direct. In general, connectionists suppose that the physical structure of the brain (that is, as a network of neurons) has a bearing on the kinds of computation that it can carry out. This is in contrast to symbolic models, which separate the procedures for carrying out a task from the physical means by which those procedures are carried out (see Fodor and Pylyshyn, 1988).

An important connectionist principle is that complex cognitive behaviour can emerge from the co-operation of simple processing elements. These elements function without a central executive to guide their actions. Connectionists also try to avoid 'hardwiring' solutions into their models. Instead, they seek to allow their models to learn complex behaviour from experience with very few in-built assumptions to guide learning. This is yet another contrast with symbolic models, which often assume that a considerable amount of knowledge is innate.

Secondly, connectionism provides a theory of learning. The interest in learning derives in part from the associationist belief that knowledge is learned from experience. The interest also derives from the idea that the brain is a general-purpose mechanism which achieves a variety of tasks, such as learning and processing, using the same basic principles. While the symbolic processing approach does sometimes address the issue of learning, there is a tendency for proponents of that approach to describe knowledge as innate and therefore they devote considerably less attention to learning than connectionists.

Thirdly, connectionist models implement computational tasks which need to be carried out in parallel and which involve the integration of many different kinds of information. This capability is also in contrast with symbolic models, which tend to operate serially and which also tend to process different types of information separately in independent modules. A fourth and closely related characteristic is that connectionism treats knowledge as an associative network. A memory trace, for instance, is represented by a pattern of activation involving a number of processing elements called nodes. Each node can participate in a number of different activation patterns, depending on its connections to other nodes. Associative memory allows for pattern completion, whereby a fragment of a memory can cue the entire memory. Associative memory also allows for pattern-matching, so that it is possible to carry out a processing task given inexact information.

Another characteristic of connectionist networks, according to Ellis and Humphreys, is that they can display rule-governed behaviour without the need for explicit rules. While it is true that connectionist networks can, without the aid of explicit rules, carry out tasks which require the application of rules in symbolic models, it is not the case, however, that connectionist networks can carry out all kinds of rule-governed behaviour. In particular, connectionist networks cannot, to date, follow recursive rules. As the previous chapter pointed out, this is a major limitation which should not be glossed over. Symbolic models, on the other hand, are designed specifically to model rule-governed behaviour.

The characteristics of connectionism outlined above will now be illustrated with reference to specific models. All these models make use of graph-theoretical representations and the associated matrix algebra. These illustrations will facilitate the description of Elman's simple recurrent network (SRN) model. Parallel processing and associative memory are illustrated with reference to the interactive activation model; learning is illustrated with reference to the three-layer backpropagation model and finally, rule-following is illustrated with reference to Elman's simple recurrent network.

4.2.1 The interactive activation model

The interactive activation model of McClelland and Rumelhart (1981) is an early connectionist model which illustrates the notions of parallel processing and associative memory found in subsequent models. The model was developed to represent the process of visual word recognition.

According to Williams (1995: 11), many researchers in the field of word recognition believe that word recognition is a process of parallel and interactive hypothesis testing. All known words and letters are simultaneously compared with the input and hypotheses about which letters are present in the input inform hypotheses about what the word is. Hypotheses about the identity of the word in turn reinforce hypotheses about which letters are present in the input. The interactive activation model attempts to capture this parallel and interactive process.

The model consists of a set of interconnected nodes arranged into layers. Nodes in the bottom most layer function as feature detectors. They detect features of letters such as vertical or horizontal lines and so on. The nodes in the next layer represent individual letters. Each feature detector node is connected to all the nodes which represent letters containing the feature detected by that feature detector node. These connections are called excitatory because when a feature detector is activated by the input, it activates all those letter nodes with which it is connected. The letter nodes are connected to each other via inhibitory connections. Thus, if the letters W and V are both activated, they will each send an inhibitory signal to the other. This inhibition captures the idea that hypotheses about which letter is in the input (W or V?) *compete* against each other.

Letter detectors, in turn, are connected via excitatory connections to word detectors, so that the activation of a given letter detector will cause the activation of all words containing that letter. Word detectors are also connected to each other via inhibitory connections, to capture the fact that hypotheses about which word is being read also compete against each other. Word detectors also send excitatory signals back to the letter detectors, so that hypotheses about the identity of the word confirm hypotheses about the identity of the letters making up that word. Processing takes place in cycles, with signals passing from letter detectors up to word detectors and down from word detectors to letter detectors. As more information becomes available in the input, certain letter and word detectors gain more activation than others. As they gain more activation, they send stronger excitatory and inhibitory signals. Given that a node sends excitatory signals to other nodes with which it is consistent, and inhibitory signals to those nodes with which it is in competition, the net result is that, after several processing cycles, a single winner emerges by virtue of a rich-get-richer and poor-get-poorer effect.

The interactive activation model shows how a set of interconnected nodes can process different kinds of information (features, letters, words) in parallel. This ability is an important characteristic of connectionist models including MacDonald and Christiansen's connectionist account of individual differences. However, the interactive activation model does not show how learning takes place. This is illustrated below with reference to the three-layer backpropagation network. The description glosses over the underlying mathematical details.

4.2.2 Backpropagation networks

An easy way to describe three-layer backpropagation networks is in terms of the flow of a liquid through a system of pipes. Imagine two containers with two pipes leading downwards from the bottom of each. Imagine two other containers which receive liquid poured down from each of the two top containers through two pipes, one pipe from each top container. The width of each pipe determines how much of the liquid poured into a top container will flow into a bottom container. By fixing the widths of all the pipes, it is possible to determine the amount of water which will flow into each of the bottom containers when a certain amount of water is poured into the top containers. This amount is obtained by adding the products of the quantity of liquid poured down each of the top containers by the corresponding pipe width. The widths of all the pipes therefore provide a way of transforming a given input into a given output.

Many aspects of language processing can be described in terms of the transformation of inputs into outputs. For instance, inflecting a word for tense can be thought of as transforming an input – the bare form – into an output – the inflected form. Similarly, assigning case roles to sentence constituents involves transforming an input – for example, a noun in preverbal position – into an output – for example, the case role 'agent'. This transformation can be carried out by a set of input nodes that are connected to a set of output nodes. These connections are associated with weights, analogous to the width of each pipe in the illustration above. Linguistic tasks such as inflecting words for tense or assigning case roles, for instance, can therefore be carried out by having an appropriate arrangement of nodes with appropriately weighted connections. By the same token, learning to perform such linguistic tasks can be thought of as learning to arrive at an appropriate set of weights for the connections between nodes. This is what backpropagation involves.

Three-layer backpropagation networks have three layers of nodes: an input layer, a 'hidden layer' and an output layer. The hidden layer is necessary because some problems require an intermediate representation in between the input and the output. Each node in the input layer is connected to each node in the hidden layer, and each node in the hidden layer is connected to each node in the output layer. Such an arrangement can be 'trained' to carry out certain tasks. Training involves devising a set of input patterns and a set of output patterns. Each input–output pair of patterns represents a pairing of some linguistic input with its output, for example, a bare verb form with its inflected form. The weights of the connections between inputs and outputs are initially random. When an input pattern is presented to the input layer, signals are sent to the hidden layer. Each hidden layer node sums up its inputs by adding the products of each input and the weight of the connection through which that input is transmitted. The sum of these products is called the activation value of that node. Each hidden layer node then transmits its activation value to each of the output nodes with which it is connected.

Each output node sums up the inputs from each hidden layer node to arrive at an output value. This output value is then compared against the ideal output pattern for the input pattern that was presented to the input layer in order to produce an error value. If the output value is greater than the target output, then the weight of each incoming connection is decreased by a small amount. The weight of the connection between each hidden unit and each input unit is also reduced by the same small amount. Thus, the error is said to be propagated backwards. The process is repeated for each output unit. A different input pattern is then presented to the input layer and the cycle of activation and back-propagation of error is repeated. When all the input patterns have been presented to the input layer, the cycle begins again with the first input pattern and so on until the level of error is reduced to a minimum. The end result is that a set of weights is obtained which enables each input to be transformed into a given output with the minimum possible error. The network is then said to have learned the task.

Some points to note here are that learning is slow and incremental. Learning also involves feedback (when the output pattern is compared against the target pattern). Further, learning is sensitive to the frequency with which a given input pattern is presented to the input layer. This is simply because the more an input pattern is presented during training, the greater the number of opportunities the network has to adjust itself in order to associate that input with an output as accurately as possible. Finally, it is important to note that learning also involves a degree of abstraction. As noted earlier, the hidden units form an intermediate representation of the input. This intermediate representation usually takes the form of a grouping of inputs that are similar with respect to certain features. As described in the next paragraph, this categorisation can be extremely fine-grained.

Connectionists sometimes represent what goes on inside a network abstractly in terms of a state space. The number of units in a hidden layer and the range of activation values which each unit can take define a state space. For instance, two nodes, each of which can take the values 1 or 0 define a space with four states, one state for each combination of values. Models can often have 20 or more hidden units. Each unit often takes continuous values between 1 and 0. Given six-digit precision, each unit might take one million possible activation values. A small number of units can therefore define a very large number of states. Each input learned by the network can be represented by a single state which is defined by the activation values of all the hidden units. Relationships between units are mirrored in relationships between their corresponding states. For instance, similar inputs are represented by states that are located more closely together in state space. When this space is examined using statistical techniques such as principal components analysis, it often possesses an intricate hierarchical organisation (see St John and McClelland, 1990).

Three-layer backpropagation networks can carry out linguistic tasks that have often been thought to be the exclusive province of rules. As indicated above, they can learn past-tense formation, case role assignment in simple clauses and even simple phrase structure grammars. However, Fodor and Pylyshyn (1988) argue that there are certain rule-governed operations which cannot be carried out by such networks. In particular, they argue that connectionist networks cannot follow recursive rules. This argument is part of a more general argument to the effect that connectionist models cannot handle structured representations. The next section describes the connectionist response to this criticism, in the form of simple recurrent networks. These networks can actually process sentences with a limited degree of self-embedding and connectionists sometimes regard their success as a refutation of Fodor and Pylyshyn's argument. However, it will be argued that the way simple recurrent networks process self-embedded sentences simply serves to demonstrate the truth of the argument that connectionist networks (to date, at least) cannot follow recursive rules.

4.2.3 Simple recurrent networks

Simple recurrent networks (Elman, 1992) are a variation of the three-layer backpropagation network with the added capacity to learn and process sequences of input patterns. As will be shown presently, the ability to learn sequences is important for the representation of syntactic structures that are more complex than the simple clause. Whereas the three-layer network consists of just three layers of nodes, the simple recurrent network has a fourth layer, called a context layer. This layer stores the pattern of hidden unit activations from one processing cycle so that they can be added onto the hidden unit activations in the next processing cycle. Thus, information about previous processing cycles is retained over several subsequent processing cycles. The context layer therefore provides the network with a short-term memory (STM) of what it has done in the recent past. The model is therefore essentially a finite state machine with memory. The memory is important for the processing of complex syntactic structures which involve long-distance dependencies. For instance, a sentence containing embedded material has a long-distance dependency between the matrix subject and the matrix verb. The context layer makes it possible to retain information about the subject verb, such as agreement information for instance, and this makes it possible to check whether or not the matrix verb agrees with the matrix subject. It is important to note, however, that while the context layer provides a kind of short-term or working memory, it does not incorporate the kinds of control mechanisms associated with this component of human memory.

Elman trained his network on a corpus of sentences, or training set, generated by a simple phrase structure grammar of English. The task of the network during training was to predict the next word in the sentence. Training was successful. The network could make correct predictions about number agreement between nouns and verbs at up to two degrees of self-embedding. When the representational space was examined, it was found that the network had induced a fine-grained hierarchical category structure. Lexical items were categorised into verbs and nouns. The verbs were subcategorised into different transitivity types. The nouns were categorised according to their animacy features. The animate nouns were further divided into human versus animal terms. The hierarchical organisation even extended to single words, so that each word occupied a unique region of the representational space. The region corresponding to each word was further subdivided according to the different contexts in which each word appeared. For instance, there was a region of the representational space which corresponded to the word 'John'. This region was subdivided into regions which corresponded to uses of the word 'John' as a subject or as an object of a verb.

The analysis showed that simple sentences were represented as trajectories across the representational space. Embedded sentences were represented as spiral-like trajectories. With increasing depth of self-embedding, the spiral became tighter, and a point was reached when the system ran out of states which could be used to distinguish the different levels of self-embedding from each other. Interference between neighbouring states therefore placed a limit on the level of self-embedding which could be modelled. In another study, Werkerly and Elman (1992) found that SRN performance on centre-embedded sentences improved when semantic constraints were imposed on subject–verb pairings. The effect of the constraints was apparently to make representations at each level of embedding more distinct from each other so that relevant information could be better preserved. This is an interesting finding, which relates to the idea, discussed in Chapter 5, that limitations in working memory are not due so much to insufficient space as to interference between representations which are not sufficiently distinct from each other.

Further simulations by Christiansen and Chater (1999) led to the conclusion that SRNs cannot process sentences with more than three degrees of self-embedding. However, neither Elman nor Christiansen and Chater describe this limited performance as a problem for connectionism. They argue instead that the performance of SRNs on self-embedded sentences mirrors that of human subjects. They argue that both human beings and SRNs process self-embedded sentences without the use of recursive rules, therefore Chomsky and Fodor and Pylyshyn's arguments are invalid. This argument is identical to the one made by Reich (1969). However, this argument is problematic because it is not consistent with the experimental data reviewed in Chapter 3. That data suggested that subjects can learn to understand sentences with up to five levels of self-embedding and to produce sentences with up to seven levels of self-embedding. This behaviour does not mirror that of SRNs, which treat self-embedded sentences simply as sequences. The degree of self-embedding supported by SRNs is limited because sequences become less predictable the longer they become. Rules, however, can enable an individual to process an arbitrary level of self-embedding, without having to memorise sequences, so long as some kind of memory aid is available.

4.2.4 Summary

Chapter 2 showed that Hockett's finite state model is rooted in associationism and a psychological interpretation of information theory. It identifies language with Saussure's parole. Knowledge of language is described as a systematisation of linguistic experience in the form of a finite state machine with a limited capacity for generalisation to new instances. The model treats sentence representations as unique trajectories through state space. It focuses on long-term memory (LTM) processes and treats language processing exclusively in terms of the laws of contiguity and frequency. The model does not incorporate mechanisms for the voluntary control of cognitive processes. It does not support recursive rules and it appears to underestimate the human potential for language.

This description applies equally to current connectionist models. While such models do differ in significant respects from Hockett's model, the fundamentals are pretty much the same. Connectionist models are associationist in character and treat language largely in terms of the laws of contiguity and frequency. Knowledge of language is regarded a systematisation of a corpus ('the training set') which has a limited capacity for generalisation to new instances. Sentence representations are described as unique trajectories in state space. Connectionist models focus on LTM processes and do not provide mechanisms for the active control of cognitive processes. Like Hockett's model, connectionist networks do not currently support the use of recursive rules and appear to underestimate the human potential for language.

4.3 Symbolic processing

The symbolic approach is based on what is sometimes called the 'computer metaphor of mind'. In terms of this metaphor, various aspects of mental organisation are made to correspond to components of computer architecture. For instance, attention is likened to a central processor, STM to RAM (temporary computer memory) and LTM to disk storage in computers. Perhaps, the most significant analogy is the one which likens human language to computer languages. Computer languages are highly regular: they consist of a clearly defined set of rules and a clearly defined set of symbols to which those rules apply. An important property of the sorts of representations employed by computers is that they are compositional and it is thought that human languages are also compositional:

> [Mental representations have] a combinatorial syntax and semantics, in which (a) there is a distinction between structurally atomic and structurally molecular representations; (b) structurally molecular representations have syntactical constituents that are themselves either structurally molecular or structurally atomic; and (c) the semantic content of a (molecular) representation is a function of the

> semantic contents of its syntactic parts, together with its constituent structure (Fodor and Pylyshyn, 1988: 12).

According to Fodor and Pylyshyn, compositionality gives rise to three key qualities of language, which they refer to as productivity, structure-sensitivity and systematicity. These qualities are described below.

4.3.1 Productivity, structure-sensitivity and systematicity

The quality of productivity is reflected in recursive rules of grammar. These rules allow human beings to understand and produce novel sentences. In self-embedding, for instance, reapplication of the same rule allows a potentially infinite number of instances of the same constituent to be nested within each other. Structure-sensitivity is manifested in the fact that the syntactic structure of an expression, and not just its content, determines how the expression is to be interpreted. For instance, 'John loves Mary' has a different interpretation from 'Mary loves John' arising from the syntactic differences between the two expressions. Finally, systematicity is manifested in the fact that someone who understands 'John loves Mary' can also understand 'Mary loves John'. This understanding is based on recognition of the fact that 'John' and 'Mary' are tokens of the same type, that is, noun. Systematicity arises because grammatical rules are defined in terms of types rather than tokens.

4.3.2 The competence–performance distinction

Another correspondence which Fodor and Pylyshyn see between humans and computers is a strict separation of knowledge from computational resources. According to Fodor and Pylyshyn (1988), there is a clear distinction between the knowledge required to do a task and the computational resources needed to carry out that task. Thus there is a clear distinction between knowledge and memory. For instance, during the processing of self-embedded sentences, a representation of the sentence is constructed by making copies of grammatical symbols and concatenating those symbols in working memory. If there is insufficient memory, then comprehension may fail. However, this failure does not detract from the abstract ability of the system to carry out the task: if more memory were available, then it would be possible to construct a representation of the sentence. In their own words:

> [...] There are a number of considerations which suggest that, despite *ipso facto* constraints on performance, ones knowledge of ones language

> supports an unbounded productive capacity in much the same way that ones knowledge of addition supports an unbounded number of sums. Among these considerations are, for example, the fact that a speaker/hearer's performance can be improved by relaxing time constraints, increasing motivation, or supplying pencil and paper. It seems very natural to treat such manipulations as affecting the transient state of the speaker's memory and attention rather than what he knows about – or how he represents – his language (34).

Fodor and Pylyshyn therefore attribute the difficulty of processing self-embedded and other complex sentences to insufficient computational resources, rather than in terms of insufficient grammatical knowledge. In addition to making a distinction between knowledge and computational resources, Fodor (1983) also proposed that the mind is modular, and that there are different modules for carrying out different cognitive tasks. Sentence comprehension, for instance, is thought to involve a module for carrying out syntactic analysis. The output of this module is then passed on to another module for further processing. Fodor's ideas inform many rule-based models of sentence comprehension which differ considerably in the way they implement these ideas. Key assumptions often made by such models, however, are that (a) there is a syntactic module for pure syntactic processing; (b) this module possesses a generative grammar of some sort; (c) syntactic processing is constrained by the availability of working memory and (d) there are supervisory mechanisms which provide overall control for the process of constructing syntactic representations. Gibson and Thomas' (1999) model makes all these assumptions. The description of this model below therefore provides a good basis for the discussion of rule-based models of individual differences in Chapter 5.

4.4 The Gibson and Thomas (1999) model

As noted above, it is widely assumed that sentence comprehension is constrained by working-memory capacity (for instance, Frazier, 1987; Just and Carpenter, 1992; Lewis, 1996). Gibson and Thomas (1999) present a metric which quantifies working-memory requirements and enables specific predictions to be made about a type of processing error called *structural forgetting*. One symptom of this error is the failure of subjects to notice an ungrammaticality caused by deleting a constituent. Gibson *et al.* attribute the failure to a memory management routine which is triggered whenever sentence complexity threatens to overwhelm working-memory capacity. The routine maintains memory expenditure below a fixed threshold by deleting syntactic nodes which make excessive memory demands.

Gibson *et al.* use their metric to account for the seeming obliviousness of subjects to an ungrammaticality in doubly nested sentences with one VP missing, for example,

1. [The patient [*the nurse* [the clinic had hired] *missing VP*] met Jack].

Gibson *et al.* propose that the memory requirements of this sentence exceed memory capacity by the time the word 'clinic' is encountered. As a result, an *S* node is deleted from the parse tree. The result of the operation is the singly nested sentence:

2. [The patient [the clinic had hired] met Jack].

Because subjects retain only the altered representation for further processing, the missing VP in (1) and its subject NP are forgotten and the sentence is judged to be grammatically acceptable.

An outline will now be given of (a) the complexity metric which quantifies memory load and (b) the 'high memory cost pruning hypothesis', which predicts exactly where in the parse tree 'pruning' must take place.

4.4.1 The syntactic complexity metric

The metric is based on five assumptions. The first assumption, referred to as 'the clause-based closure hypothesis', is that syntactic processing is clause-based. It is assumed that a completed clause is shunted out of working memory and therefore it imposes no memory cost on subsequent processing. After a completed matrix clause is shunted out of working memory, the next clause becomes the matrix clause and will similarly impose no memory requirements after it too has been processed. Thus, an arbitrarily recursive right-branching structure can be parsed using a limited memory.

The second assumption is that, at any point during the construction of a parse tree, the syntactic processor predicts the syntactic heads which are required to complete the sentence grammatically. Each predicted head is associated with a single unit of memory, referred to as M.

The third assumption is that there is a memory cost associated with each new discourse referent that is processed after a specific prediction has been made, symbolised by (n). Therefore, if one head has been predicted and one discourse referent has been processed since that prediction was made, then the memory cost is 1M(1) memory units, where 1M is the number of predicted heads and (n = 1) is the number of discourse referents processed since that prediction was made.

The fourth assumption is that at least two heads are required for a sentence – a noun for the subject and a verb for the predicate. The fifth assumption is that there is no cost associated with the prediction of the matrix verb.

With these assumptions in mind, the calculation of memory requirements for the following pair of sentences proceeds as follows:

3. The professor who_i the scientist collaborated with e_i had advised the student who_j e_j copied the article.
4. The student who_j the professor who_i the scientist collaborated with e_i had advised e_j copied the article.

Gibson *et al.* state that the point of highest complexity in (3) is at 'the scientist', where three heads are required to complete the sentence grammatically: (a) the matrix verb; (b) the verb which heads the relative clause predicate; and (c) the NP empty category (e_i) to be co-indexed with the relative clause pronoun 'who'. The memory cost is calculated as follows: (a) for the matrix verb there is no memory cost; (b) for the verb which heads the relative clause predicate, M = 1. This prediction is first made at the point marked by 'who'. At this point, n = 1 because one new discourse referent 'the scientist' has been processed since the prediction for the relative clause was made (at the point marked by 'who'). Thus the memory cost is M(1); (c) An identical calculation is carried out for (c): the empty NP category was predicted at 'who' (M = 1) and one new discourse referent has been processed since this prediction was made (therefore n = 1). Therefore the memory cost for (c) is also M(1). All together, the total memory cost for (3) at 'the scientist' is M(1) + M(1) memory units = 2M(1).

In (4), the memory cost at 'the scientist' is higher than in (3). At this point, five rather than three heads are required to complete the sentence: (i) the matrix verb; (ii) the verb of the first relative clause; (iii) the verb of the second relative clause; (iv) an NP empty category to be co-indexed with relative pronoun of the first relative clause and (v) an empty NP category to be co-indexed with the relative pronoun of the second relative clause. For (i) there is no cost. For (ii) the cost is 1M and n = 2 because two discourse referents have been processed since the prediction was first made, therefore the cost is M(2). For (iii) the cost is 1M and n = 1 because one discourse referent has been processed since the prediction was first made, therefore the cost is M(1). For (iv) the cost is 1M and n = 2 because two discourse referents have been processed since the prediction was first made, therefore the cost is M(2). For (v) the cost is 1M and n = 1 because one discourse referent has been processed since the prediction was first made, therefore the cost of (v) is M(1). The total cost for (4) is therefore M(2) + M(2) + M(1) + M(1) = 2M(2) + 2M(1). The higher memory cost of (4) relative to (3) is said to account for the greater complexity of (4).

The high memory cost pruning hypothesis states that: 'At points of high memory complexity, forget the syntactic prediction(s) associated with the most memory load.' (231). In (4), predictions relating to the first relative clause are associated with the highest memory load – 2M(2) – therefore this clause is 'pruned', resulting in structural forgetting. The sentence that remains after structural forgetting would therefore be:

5. The student who_j the scientist collaborated with e_i e_j copied the article.

Gibson *et al.* do not state the precise memory cost at which structural forgetting occurs. Presumably it lies between 2M(1) (the maximum memory load for (3)) and 2M(2) + 2M(1) MUs (the maximum memory load for (4)).

4.4.2 Experimental support

Gibson *et al.*'s theory receives some support from their (1999) experiment quite well. In their experiment, subjects were asked to rate the complexity of doubly nested sentences under four conditions: (a) all three VPs present; (b) missing VP1; (c) missing VP2 and (d) missing VP3, for example (the purpose of the italics will be explained presently)

1. All VPs present: The ancient manuscript that the graduate student who the new card catalogue had confused a great deal was studying in the library was missing a page.
2. Missing VP1: The ancient manuscript that the graduate student who *the new card catalogue was studying* in the library was missing a page.
3. Missing VP2: The ancient manuscript that the graduate student who the new card catalogue had confused a great deal was missing a page.
4. Missing VP3: *The ancient manuscript* that the graduate student who the new card catalogue had confused a great deal *was studying in the library*.

The complexity ratings were (1)<(3)<((2) = (4)). (1) has the greatest structural complexity so its 'least complex' rating is seriously problematic. The authors put the anomalous result down to a task effect but the validity of the complexity-rating task is put into doubt. The remaining pattern of results is consistent with the proposal that the internal representation of missing VP2 sentences is simpler than that of missing VP1 or missing VP3.

4.4.3 Problems interpreting the results

There is alternative account for the data which makes no reference to node deletion and accommodates the anomalous 'least complex' rating for the all three VPs present condition. This account is based on evidence that many native English subjects cannot understand doubly nested structures in the absence of semantic constraints (see Blaubergs and Braine, 1976; Blumenthal, 1966; Freedle and Craun, 1970; Miller and Isard, 1964; Powell and Peters, 1973; Stolz, 1967; see also Chapter 3 for a review of these studies). There is also evidence that subjects can use processing strategies which do not involve the construction of parse trees (see Townsend and Bever, 2001). For instance, a canonical sentence strategy allows subjects to interpret an *NP VP* sequence as a clause. Subjects might also use a relative clause strategy which treats the first NP and last VP of a complex sentence as the matrix clause. It is shown below how Gibson *et al.*'s results can be explained in terms of a combination of semantic and word order heuristics.

Firstly, plausibility and the syntactic position of the omitted VP are confounded in Gibson *et al.*'s stimuli. Gibson *et al.* used verbs with strong semantic/pragmatic selectional restrictions, so that only one preceding NP was a plausible subject for each verb. In (1) each VP can be associated with a plausible NP subject and therefore the sentence poses the least problems for semantically based comprehension; hence the 'least complex' rating. In (2), however, the canonical sentence strategy of interpreting an *NP VP* sequence as a single clause clashes with the semantic incongruity of the subject and predicate pair 'who *the new catalogue was studying...*', hence the 'most complex' rating. In (3), on the other hand, there is no semantic incongruity. Although 'student' lacks a predicate, it does function as the object of a VP, so that every NP in the sentence can be associated felicitously, albeit incompletely, with a VP. Hence (3)<(2). In (4), the matrix NP and VP are semantically mismatched '*The ancient manuscript* (embedded material) *was studying in the library*'. In this case, the use of a relative clause strategy to interpret the first NP and last VP as belonging to the matrix clause clashes with the semantic incongruity of the NP and VP. Hence ratings of equal complexity to (2) and (4).

4.4.4 Summary

Chapter 2 also showed that Chomsky's generative model arises from a rationalist view of psychology and a computer metaphor of mind. It treats knowledge of language as a deductive system, which takes the form of an innate generative grammar. The model treats sentence representations as symbol structures which record the derivational history of the sentence in the form of phrase structure trees. It requires STM to construct temporary, intermediate representations (Chomsky, 1965). The model supports recursive rules. However, in view of the experimental evidence reported in Chapter 3, it appears to overestimate the actual linguistic abilities of language users by hardwiring the ability to use recursive rules.

These observations apply equally to current rule-based models, of which the one proposed by Gibson and Thomas (1999) is a prime example. These models adopt a rationalist view of psychology. They treat knowledge of language as a deductive system which generates phrase structure representations of sentences. Chapter 2 presented an argument to the effect that rule-based models do not address the problem of overgeneration, the core–periphery distinction and the likely consequences of individual differences in lexical knowledge for syntactic processing. Rule-based models generally assume that native speakers possess a uniform competence which is only indirectly reflected in performance. These models explicitly invoke the notion of working memory and some of them given detailed descriptions of the control processes involved in the construction of linguistic representations. In view of the experimental evidence presented in Chapter 3, these models also appear to overestimate actual human linguistic abilities by hardwiring recursive rule systems.

4.5 Conclusion

While this chapter has shown that the detailed operation of current models is different from the earlier ones, it has also shown that current experience-based and rule-based models make underlying assumptions about linguistic knowledge that are identical to those made by Hockett and Chomsky, respectively. The experimental work reviewed in Chapter 3 indicated that these assumptions yield only partial descriptions of what language users know about language.

It may be useful, at this point to revisit the experimental findings reviewed in Chapter 3.

Those findings indicated that (a) language has a statistical structure and language users are sensitive to it; (b) language users vary in their sensitivity to the statistical structure of language; (c) knowledge of the statistical structure of language helps language users reduce limitations in performance usually attributed to intrinsic limitations of the cognitive machinery; (d) it is not clear whether or not language users possess phrase structure rules; (e) language users can employ recursive rules but they differ in their ability to do so; (f) language users can be trained to use recursive rules and (g) language users differ in syntactic competence.

Elman's connectionist model captures the statistical aspects of linguistic knowledge. However, it cannot explain the ability of language users to apply recursive rules to an arbitrary degree. Gibson and Thomas' model cannot explain the statistical aspects of language use. It does capture the ability of language users to apply recursive rules. However, it hardwires the ability to use recursive rules into the language processing architecture and therefore cannot explain the fact that some language users can only use recursive rules after training. In addition, the model does not address the problems of overgeneration, undergeneration and the possible impact of individual differences on lexical knowledge in syntactic processing.

It is therefore fair to say that current models do not address the early experimental findings. Given that current models of sentence comprehension share the same assumptions as the earlier models, the early experimental findings are just as problematic for current models as they were for the earlier ones. This conclusion also applies to current theories of individual differences, which are either based on or closely related to the models described in this chapter. Current models of individual differences are discussed in the next chapter.

5
Current Theories of Individual Differences

5.1 Introduction

This chapter describes three current models of individual differences in the comprehension of complex sentences: Just and Carpenter (1992), Caplan and Waters (1999) and Christiansen and MacDonald (2000). It is argued that these models inherit the limitations of rule-based and connectionist models described in Chapter 4 and consequently cannot fully explain all the data on individual differences. The chapter then describes two hybrid models of sentence comprehension that attempt to integrate elements of connectionist and symbolic processing – Miikkulainen (1996) and Townsend and Bever (2001). It is argued that these models too cannot explain the individual differences described in Chapter 3. The chapter concludes by proposing that an account of individual differences needs to integrate the long-term memory (LTM) processes which underlie associative learning with the conscious control processes that underlie rule-governed behaviour and to allow individuals to vary in terms of both processes.

The chapter is divided into five sections. Each section describes, in turn, the models of Just and Carpenter (1992), Caplan and Waters (1999), MacDonald and Christiansen (2002), Miikkulainen (1996) and Townsend and Bever (2001).

5.2 Just and Carpenter's model

The description of Just and Carpenter's theory is organised into five sub-sections. The first sub-section describes their basic definition of working memory. The second relates the notion of working memory to the notion of activation. The third describes various memory management schemes which control the dynamic allocation of memory during processing. The fourth describes the effects of individual differences in memory capacity on comprehension. The final sub-section considers Just and Carpenter's approach to learning.

5.2.1 Definition of working memory

The point of departure for Just and Carpenter's theory is Baddeley's model of working memory. Baddeley and Hitch (1974) argued that the concept of short-term memory (STM) needed to be replaced by the concept of working memory. The replacement was necessary because STM is restricted to storage of information and does not include the manipulation of information. The concept of working memory, on the other hand, includes both the temporary storage and manipulation of information.

Working memory as envisaged by Baddeley consists of three components: a phonological loop, a visuo-spatial sketch pad and a central executive. The phonological loop and sketch pad perform storage functions while the central executive performs processing functions. Just and Carpenter state that their notion of working memory is equivalent to the central executive: 'The working memory in our theory corresponds approximately to the part of the central executive in Baddeley's theory that deals with language comprehension' (123). Just and Carpenter therefore regard both processing and storage as functions of one undivided system. In their view, working memory carries out symbolic manipulations that are:

> at the heart of human thinking – such operations as comparison, retrieval, and logical and numerical operations. Of particular relevance are the processes that perform language comprehension. These processes, in combination with the storage resources, constitute working memory for language (123).

Just and Carpenter restrict their discussion of working memory to purely verbal processes: 'In our theory, working memory for language refers to a set of processes and resources that perform language comprehension' (123):

> Working memory plays a critical role in storing the intermediate and final products of a reader's or listener's computations as he or she constructs and integrates ideas from the stream of successive words in a text or spoken discourse. In addition to its role in storage, working

> memory can also be viewed as the pool of operational resources that perform the symbolic computations and thereby generate the intermediate and final products (122).

5.2.2 The concept of activation

The storage and processing functions of working memory depend on the amount of available activation. In the theory, representational elements such as words, phrases, propositions, grammatical structures and thematic structures, and so on are each associated with a certain level of activation. These representational elements become activated during comprehension whenever they are encoded from texts or speech; generated by a computation or retrieved from LTM. For instance, encountering a subject NP gives rise to the expectation of a VP. Thus, the NP is said to cause the VP to become activated. An element is said to be in working memory as long as its activation level exceeds a certain threshold. And as long as an element is in working memory, it can be manipulated by one process or another.

If the number of activated elements in working memory exceeds the total available amount of activation, then activation may be deallocated from some elements, and those elements are then forgotten. For instance, words which occur earlier in a sentence may be forgotten if the sentence is too long. Comprehension will fail if the words which are needed later in processing have been forgotten. Just as a number of representational elements may be kept active in working memory, a number of processes can run concurrently. For instance, while the appearance of an NP is causing a VP to become activated, other processes, involving syntactic, semantic and pragmatic aspects of the sentence take place concurrently. When the number of concurrent processes exceeds the total amount of available activation, then some processes may be scaled down by making them run more slowly.

5.2.3 Memory management mechanisms

Since memory requirements in normal comprehension often exceed available activation, certain memory management schemes are in place to allocate and deallocate memory in such a way as to prevent comprehension failure: 'these mechanisms [...] minimise the demands on storage [and] keep the overall demands on working memory manageable, even when an extended text is being processed.' (124). Just and Carpenter explain the operation of memory management schemes as follows:

> Because a text can contain an indefinitely large number of sentences whose storage can eventually consume a finite capacity, there must be countervailing mechanisms that reduce the storage demands. Some of these mechanisms selectively retain representations of only the most recent and most central clauses in an activated form, while dampening the activation level of other propositions from earlier sentences. Moreover analogous mechanisms may selectively retain only the most relevant aspects of world knowledge in an activated form, while dampening the activation level of other knowledge that might be initially activated by the reading of the text (124).

For instance, context is said to facilitate text processing by 'preactivating some concepts, relations and schemas relevant to its comprehension' (124). Working memory therefore depends on the existence of memory management mechanisms to ensure that comprehension does not fail.

5.2.4 Individual differences

The central idea in Just and Carpenter's (1992) article is that individual differences in working-memory capacity affect individual differences in comprehension. There are three ways in which individual differences in working memory affect the nature of comprehension. Firstly, individuals with smaller working-memory capacities are more susceptible to forgetting. They are therefore vulnerable to comprehension failure than individuals with larger working-memory capacities. Secondly, individuals with smaller working-memory capacities are also liable to have lower processing speeds. As explained above, one way of reducing demands on activation is to reduce the speed of a process. Thirdly, individuals with smaller working-memory capacities can run fewer processes concurrently. This means that such individuals may be unable to carry out interactive processing when memory demands are high:

> According to our view, people with small working memories for language may not have the capacity to entertain (keep activated and propagate additional activation from) nonsyntactic information during the syntactic computations, or at least to the degree that the nonsyntactic information can influence the syntactic processing (126).

Just and Carpenter therefore explain individual differences in comprehension in terms of individual differences in working-memory capacity and not in terms of grammatical knowledge. Just and Carpenter measure working memory in terms of Daneman and Carpenter's (1980) reading span test. In the test, subjects are given sets of sentences which they must read while maintaining the final word in memory. The size of an individual's working-memory capacity is given by the size of the set from which he or she can recall all the final words. Just and Carpenter report that reading ability is correlated with reading span, a finding which they interpret as support for their theory.

5.2.5 Learning

Just and Carpenter do not give much attention to the possibility that individuals may differ in the linguistic knowledge necessary for comprehension. They restrict their discussion of individual differences in knowledge to individual differences in the degree to which certain processes are practised. There is an implicit assumption that the same processes are present in all individuals:

> [...] changes in the efficiency of a process are often assumed to result from practice or some instructional intervention. Indeed, extensive practice in several simple tasks, such as Stroop-like tasks, induce large changes in the speed of responding that are typically interpreted in terms of changes of efficiency of underlying processes (145).

Just and Carpenter mention that some individual differences in reading performance are due to differences in practice:

> Intensive practice in reading might similarly induce greater efficiency in some component processes of comprehension; the time spent in out-of-school reading is correlated with reading skill in fifth-grade students, accounting for approximately 9% of the variance in one study (145).

However, there is no mention of individual differences in syntactic knowledge. Just and Carpenter therefore adopt a rule-based conception of language comprehension, whereby individual differences in performance are attributed primarily to differences in memory capacity.

5.2.6 Relationship to known patterns of individual differences

Just and Carpenter (1992) do not address findings of individual differences in phonology, morphology or the lexicon (see Chapter 3). Individual differences in these areas are particularly problematic for the theory because they exist at levels of structural complexity that are well below the levels at which putative working-memory effects are reported. Secondly, Just and Carpenter assume that native speakers possess phrase structure grammars but provide no justification for this assumption. As pointed out in Chapter 2, no conclusive evidence was ever produced to show that native speakers possess generative grammars. In addition, Just and Carpenter do not address the problems of overgeneration or undergeneration. Thirdly, Just and Carpenter do not address the possibility of individual differences in lexical knowledge and the logical impact of such differences in syntactic processing.

There is a fourth problem with Just and Carpenter's model. As described above, Just and Carpenter describe the use of semantics in syntactic comprehension as something which requires high working-memory capacity. In their account, only individuals with high working-memory capacity can use semantic constraints during syntactic comprehension. It is therefore hard to explain why, in terms of the theory, some individuals can only comprehend certain types of sentences when semantic constraints are provided, as indicated by the experiments carried out by Stolz (1967) and Powell and Peters (1973). A final problem for Just and Carpenter's model is that it is difficult to say why linguistic training should result in improved comprehension of self-embedded sentences (Freedle and Craun, 1970; Powell and Peters, 1973; Stolz, 1967). If working-memory capacity is an inherent dimension of individual differences and the sole constraint on sentence processing, then it is not clear why training should improve performance, other than to make an existing process more efficient.

5.3 Caplan and Waters' model

Caplan and Waters (1999) criticise Just and Carpenter's theory on the basis that they have not been able to replicate Just and Carpenter's results. They report a number of experiments in which they found no effect of working-memory capacity on syntactic processing. They report, however, that they find an effect of working-memory capacity on the ability to use the propositional information generated by syntactic processing. Caplan and Waters then propose their own theory of individual differences. Their proposals are not as detailed as those of Just and Carpenter (1992), and therefore the description of their approach below will be much briefer.

5.3.1 Controlled and automatic processes

The theory proposed by Caplan and Waters is based on a distinction between controlled and automatic processes (Schiffrin and Schneider, 1977).

In Schiffrin and Schneider's theory, automatic processes do not require conscious attention and are therefore said not to require processing resources (another term for working memory). Conscious processes, however, are thought to require processing resources. Caplan and Waters combine Schiffrin and Schneider's theory with Fodor's (1983) modularity hypothesis. They equate automatic processing with the kinds of processes carried out by Fodor's modules. Caplan and Waters propose that syntactic processing, which they call interpretive processing, is an automatic process which is carried out by a syntactic module. However, subsequent processing, which involves manipulating the propositional content of sentences, is non-automatic. According to Caplan and Waters, individuals differ in working memory for processes involving the use of propositional information, such as remembering the content of a sentence, reasoning about that content or using that content to plan actions. In other words, individuals differ in the working memory needed for non-automatic processes.

5.3.2 Modularity versus learned skills

It is difficult to evaluate Caplan and Waters' theory for the following reasons. Firstly, six of the criteria which Fodor (1983) uses to define modules closely resemble characteristics of learned skills (see Bates *et al.*, 1988: 23–24). Caplan and Waters are themselves uncertain whether the specialised working memory for syntactic processing is innate or is the result of learning. This uncertainty is problematic because the sorts of evidence which can be used to support the existence of a syntactic module are ambiguous between innate specification and learning. A second and closely related problem is that evidence of a dissociation between syntactic and propositional processing could arise from a dissociation between overlearned and novel aspects of a comprehension task. This last problem can be illustrated below with reference to the processing of self-embedded sentences and with effects of expertise on text processing.

Chapter 2 reviewed evidence from Freedle and Craun (1970), Powell and Peters (1973) and Stolz (1967) which showed individual differences in the ability to understand self-embedded sentences. Presumably, Caplan and Waters would explain these differences by saying that all subjects could process the test sentences syntactically, but that subjects with low working-memory capacity had difficulty processing the propositional content of those sentences. However, both Stolz and Powell and Peters found that the poor performers could understand the sentences when given semantic constraints. It is not clear how Caplan and Waters' theory can account for this result. If subjects with low working memory have difficulty with propositional processing, then they should have displayed such difficulties whether the sentences were semantically plausible or semantically neutral.

It is possible to explain this result by showing that automaticity can pertain equally to either syntax or semantics and, by the same token, that novelty can pertain equally to either syntax or semantics.

Firstly, consider the following example:

The robber arrested the policeman.

In this sentence, the syntax is highly familiar and it is reasonable to suppose that any purely syntactic measure of processing will detect no differences between native speakers of English. However, the propositional content of the sentence is novel, because robbers do not normally arrest policemen. In order to reason about the propositional content of this sentence, an individual needs to be able to construct some context in which it is plausible for a robber to arrest a policeman. It might be, for instance, that the robber has been drafted in an undercover operation and given the authority to arrest a corrupt policeman. It is reasonable to suppose that individuals might differ in their ability to construct such contexts and that such differences might be measurable in terms of processing time or some other on-line measure. Now consider the following sentence:

The honey that the bear that the hunter shot ate was made by African bees.

In this case, the syntax is complex but the propositional content is not. It is well known that hunters shoot animals such as bears and that bears eat honey. As reported in Chapter 3, Stolz (1967) found that his subjects could process self-embedded sentences if semantic constraints were provided. However, in the following sentence:

The farmer that the hunter that the fireman knew telephoned went home.

The only way to work out who did what to whom is by paying attention to the syntax. As reported in Chapter 3, Stolz found that half his subjects could not comprehend such sentences. The most likely reason for this was that his subjects were unfamiliar with the syntax of self-embedded sentences. This observation receives further support from the fact that individuals benefited from training in syntactic analysis.

Chapter 8 will also describe how individuals who are experts in a given domain tend to perform well in comprehending texts about that domain. The effect of expertise is so strong that poor readers who are experts in football comprehend texts about football better than skilled readers who are ignorant of football. A simple explanation for the effects of semantics on the processing of self-embedded sentences and the effects of expertise on text processing is that familiarity with content facilitates processing of language associated with that content. Semantically plausible self-embedded sentences are easier to process than semantically neutral ones because subjects can make use of their long-term semantic memory to match subjects to verbs. By the same token, football experts can make use of their LTM of football knowledge to understand the content of a text on football. Conversely, familiarity with a given syntactic structure may facilitate the processing of a sentence at a purely syntactic level, just as familiarity with texts may facilitate the processing of domain-neutral texts.

In sum, sentence comprehension involves a number of sub-tasks, and these include syntactic and semantic processing. The familiar tasks may be carried out automatically while the unfamiliar tasks may require non-automatic processing. There is no reason to associate automatic processing uniquely with syntax or to associate non-automatic processing uniquely with propositional processing. The reverse can be true. Caplan and Waters could therefore be confounding automaticity with syntactic processing, and conscious processing with propositional processing.

5.3.3 Innateness versus experience

Another problem is that Caplan and Waters' theory makes two different kinds of predictions depending on whether syntactic processing is learned or innate. If syntactic processing is innate, then individual differences in linguistic experience should have no bearing on purely syntactic processing. If syntactic processing is learned, on the other hand, then one would expect individual differences in experience to have a significant bearing. As it stands, therefore, the theory doesn't make clear predictions about the relationship between linguistic experience and syntactic processing. It is also worth noting that Caplan and Waters' model is subject to the criticisms that were made earlier about Just and Carpenter's model, to do with problems of the rule-based approach.

5.4 MacDonald and Christiansen's model

5.4.1 Individual differences in ambiguity resolution

MacDonald *et al.* (1992) studied the on-line processing of main verb versus reduced relative clause ambiguities, for example,

> *The experienced soldiers warned about the dangers before the midnight raid* (where 'warned' is the main verb of the sentence).
> *The experienced soldiers warned about the dangers conducted the midnight raid* (where 'warned' is the main verb of the reduced relative clause).

They found that subjects categorised as having high working memory (high span) spent more time reading the region of the sentence which contains disambiguating information compared to control sentences, whereas low span subjects did not. This result was explained by saying that the high span readers could hold two alternative interpretations of a sentence in working memory while the low span readers could only hold one interpretation in working memory. Because the high span readers had to decide between two interpretations of the sentence, they spent more time reading the disambiguating information, whereas the low span readers had only one interpretation to deal with and therefore read the ambiguous sentences as quickly as they read the unambiguous controls. However, subsequent work by MacDonald suggested that high and low span individuals differed in their knowledge of contextual information relating to the ambiguity.

5.4.2 Contextual constraints

First, MacDonald (1994) carried out some experiments which indicate that language users are sensitive to contextual lexical constraints during the processing of ambiguous sentences. A chief constraint which she has studied concerns the frequencies of various argument structures of ambiguous verbs. For instance, the sentence fragment, 'The patients heard . . . ' is compatible with at least four continuations consistent with four argument structures of the verb 'heard'. These are (a) an active transitive argument structure, as in 'The patients heard the music'; (b) an intransitive structure, as in 'The patient heard with the help of a hearing aid'; (c) a sentential complement structure, as in, 'The patient heard that the nurses were leaving'; and (d) a reduced relative structure, as in, 'The patient heard in the cafeteria was complaining'.

Another constraint studied by MacDonald concerns the presence or absence of post-ambiguity information which makes an active transitive reading of 'heard' less likely. For instance, in 'The patient heard *in the cafeteria* was complaining', the phrase 'in the cafeteria' makes an active transitive reading of the verb unlikely. Another constraint concerns pre-ambiguity information which makes some interpretations more plausible than others. For instance, in '*The music* heard', the presence of 'music' makes an active transitive interpretation less likely than a reduced relative interpretation. 'Music' is therefore a better pre-ambiguity constraint in a reduced relative construction than 'patient'.

MacDonald found that the presence of these lexical constraints affected reading times and comprehension responses. The more frequently a verb was used intransitively in everyday speech, the more difficult it was for subjects to interpret it correctly in a reduced relative construction (as measured by reading times). It was also found that ambiguity effects were smaller with good post-ambiguity constraints than with poor ones. Good pre-ambiguity constraints were also more helpful for ambiguity resolution than poor ones. MacDonald found interactions among these constraints. For instance, the largest ambiguity effects were obtained when both pre- and post-ambiguity constraints were poor.

5.4.3 Individual differences in sensitivity to contextual constraints

Perlmutter and MacDonald (1995) found individual differences in the use of such lexical constraints during sentence comprehension. Subjects were classified as having a high or low reading span as determined by Daneman and Carpenter's (1980) reading span test. Subjects were asked to read sentences like, 'The soup cooked in the pot but was not ready to eat'. The sentence is temporarily ambiguous because the verb 'cooked' can be interpreted, among other possibilities, as a main verb in an intransitive structure: 'The soup cooked in the pot but was not ready to eat.'; a main verb in a transitive structure: 'The soup cooked the vegetables.' or a reduced relative verb in an intransitive structure: 'The soup cooked in the pot was delicious.' There is a second point of ambiguity in the test sentence where one must decide between a main verb and a reduced verb interpretation. If the sentence were to continue with 'but was not ready to eat' then 'cooked' is a main verb. However, if the sentence were to continue with 'was delicious' then 'cooked' would have to be the main verb.

Subjects were presented with ambiguous and unambiguous sentences, for example, 'The soup cooked/bubbled in the pot but was not ready to eat'. It was found that, for ambiguous sentences, high span subjects spent more time reading the ambiguous regions of the sentence than the low span readers. This finding is similar to that of Cupples and Holmes, who found that less skilled readers did not show any effect of syntactic ambiguity. Another experiment by Perlmutter and MacDonald reported in the same article suggested that both high and low span subjects behaved identically if an off-line task was used. This suggests that both groups of subjects did not differ in knowledge of lexical constraints but in using such constraints during on-line comprehension. The authors suggest that the two groups differ in their sensitivity to the frequency with which the constraints occur in natural language but do not differ in linguistic knowledge per se.

5.4.4 Connectionist modelling of individual differences

MacDonald and Christiansen (2002) present data from connectionist simulations which suggest that individual differences in processing relative clause sentences are attributable to individual differences in experience rather than individual differences in working-memory capacity, as proposed by Just and Carpenter (1992). Just and Carpenter had observed that individuals differ in processing subject- versus object-relative sentences, for example, 'The reporter that attacked the senator admitted the error' (subject relative) versus 'The reporter that the senator attacked admitted the error' (object relative). Just and Carpenter found no differences between individuals categorised as having high or low working memory in the processing of subject-relative clauses. However, they found an effect of working-memory capacity during the processing of object relative sentences. Just and Carpenter proposed that the comprehension of object relatives requires greater amounts of working memory than the processing of subject relatives, and that is why individuals with low working-memory capacity found object relatives particularly difficult.

However, MacDonald and Christiansen trained SRNs on a corpus generated by a simple phrase structure grammar of English. The training set consisted of both simple sentences and object and subject relatives. The networks had to learn to predict the next word in each sentence. It was found that the performance of the networks varied depending on the amount of training they had, and that these variations corresponded to the differences reported by Just and Carpenter between high and low span individuals. Although there were equal numbers of subject and object relatives, it was found that the network performed better on subject relatives. This was explained by saying that the networks could generalise from their experience of simple sentences to subject-relative sentences. In addition, performance on object relatives improved with greater training. These results suggest that individual differences in the processing of subject and object relatives can be modelled by SRNs with different amounts of training. The results therefore support the claim that individual differences in comprehending complex sentences are due to individual differences in experience.

The full details of MacDonald and Christiansen's argument are too extensive to discuss in full here. The important point is that they provide an alternative explanation for data which has previously been used to support a working-memory-based account of individual differences. They show that variations in the amount of linguistic exposure allow an SRN to model patterns of individual differences in comprehension of complex sentences. While the SRN appears to be a good model for the pattern of individual differences studied by MacDonald and Christiansen, other considerations indicate that the model cannot generalise to other aspects of sentence comprehension. These limitations arise from the limitations of SRNs that were described in Chapter 4.

5.4.5 Limitations of the SRN model

Firstly, Miikkulainen (1996) showed that SRNs do not generalise well to novel structures. This is an important point to clarify because connectionists often state that SRNs can generalise to novel sentences. This claim is limited to cases where new lexical items are used in the same sentence structure as that on which the network has been trained. This is not the same as being able to generalise from one structure to another, as shown by Miikkulainen (see discussion later in this chapter). Secondly, SRNs cannot model, at least in their current form, the experimental data on the processing of self-embedded sentences. Chapter 3 discussed evidence which indicates that language users can be trained to produce and understand sentences with degrees of self-embedding far in excess of the maximum of the three self-embeddings which SRNs are capable of. This evidence indicated that language users could learn recursive rules, that is, rules that can be applied over and over again, indefinitely. SRNs cannot, at present, represent such rules. This is a serious limitation, for the following reasons.

To begin with, connectionism is intended to be a general theory of mind. However, the mind would be extremely limited if it were incapable of representing rules. Mathematical reasoning, for instance, would not be possible without rules. The growth of mathematical reasoning suggests that there are no limits in principle to the structural complexity of mental representations. It is also not clear how the notion of rule, as such, could arise from a mind incapable of representing rules. Secondly, there is evidence from the psychology of memory, reviewed in the next chapter, which indicates that the structural complexity of mental representations can be quite considerable, especially in the case of skilled individuals.

In summary, while MacDonald and Christiansen's results might account for individual differences in the processing of certain types of complex sentences, their account cannot be generalised to cover the experimental data on the processing of self-embedded sentences. Further, the SRNs cannot model the kinds of complex representations constructed by skilled individuals. This limitation arises in part because the SRNs make no provision for the control processes needed to deliberately construct and use representations in LTM.

The next section describes a hybrid model which overcomes this limitation by incorporating a central executive in a connectionist architecture.

5.5 Miikkulainen's model

Miikkulainen (1996) reviews the literature on SRNs and shows that these models do not generalise well to new sentence structures. Whenever SRNs are said to generalise to novel sentences, this generalisation is restricted to cases where the same structure is used with different lexical items. In his own simulations, Miikkulainen found that a network which had learned to assign case roles to sentences like, 'The girl who liked the dog saw the boy' could generalise to sentences like 'The dog who bit the girl chased the cat' but could not generalise to novel combinations of clauses, such as 'The girl who liked the dog saw the boy who chased the cat'. Miikkulainen (1996: 2) notes:

> The problem is that the current distributed neural network architectures function simply as pattern transformers and they generalise by interpolating between patterns on which they were trained. They cannot make inferences by dynamically combining processing knowledge that was previously associated to different contexts, such as processing a relative clause in a new place in an otherwise familiar sentence structure. This lack of generalisation is a serious problem, given how effortlessly people can understand sentences that they have never seen before.

5.5.1 Incorporating control processes in a connectionist network

Miikkulainen proposes a modular architecture which consists of three types of networks, an SRN which functions as a parser, a three-layer feedforward network which functions as a clause segmenter and central executive and a recursive auto-associative memory or RAAM network, which functions as a memory device. The RAAM network is a special architecture which allows input patterns to be compressed recursively in its hidden layer. The RAAM network functions like a stack, whereby items are stored and retrieved on a first-in-last-out basis. This retrieval scheme is suitable for dealing with centre-embedded sentences, where the first subject NP has to be associated with the last VP to appear in the sentence.

In Miikkulainen's modular architecture, the SRN assigns case roles to clause constituents at its output layer. If a new clause begins before the current clause is complete, as in a centre-embedded sentence, the central executive detects this and causes the subject noun of the top-level clause to be pushed onto the stack. The embedded clause is then processed as a simple clause, unless it too contains an embedded clause, in which case its subject noun is also pushed onto the stack. The central executive detects the end of an embedded clause and causes the stack to pop the top-level noun back into the parser. The parser can then associate the noun with the appropriate verb. Thus sentences with several degrees of self-embedding can be parsed successfully.

An important aspect of the performance of this architecture is that it can generalise to new sentence structures. It has this ability because the parser only deals with single clauses and it does not have to deal with relationships between different clauses. Additionally, the parser is free from the interference which would otherwise arise from having to deal with both processing and storage functions at the same time. It is important to note that Miikkulainen enables his model to process recursive structures by incorporating a central executive which ensures that processing and storage functions do not interfere with each other. The use of a complex architecture incorporating a central executive, stack and parser violates the Connectionist maxim of allowing complex organisation to emerge from the interaction of simple processing elements.

5.5.2 Application to individual differences

Miikkulainen's model was not created to explain individual differences: its architecture is specifically customised for the processing of recursive structures. This means that the model cannot be readily applied to individual differences in the processing of either self-embedded sentences or other sentence structures. One possible application of the model to individual differences in processing self-embedded sentences is to suggest that individuals have different cognitive architectures. However, it would then be difficult to explain how individuals can *learn* to process self-embedded sentences. An intriguing possibility is that learning might actually result in the development of a task-specific architecture through a reconfiguration of connections between neurons. Some thoughts along these lines will be presented in the final chapter.

The next section describes a different approach to integrating connectionist and symbolic processing.

5.6 Townsend and Bever's hybrid model

Townsend and Bever (2001) propose a hybrid model of the language user. The model is intended to capture both statistical and rule-governed aspects of language. The model therefore has two major components: a connectionist processor and a symbolic processor which implements Chomsky's (1995) Minimalist theory of language. Without going into details, the Minimalist theory characterises knowledge of language as a formal system which derives syntactic structures from a set of initial elements called a 'numeration'. The connectionist approach as already been described.

5.6.1 Integrating associations with syntactic derivation

The hybrid model combines dual processing with analysis-by-synthesis, whereby approximations to the input are successively synthesised and compared against the input until a suitable match is found. The model operates as follows. An input string is first placed in a temporary store and then subjected to preliminary analysis by the Connectionist component of the model. The preliminary analysis determines major phrases and their conceptual relationships. The output from this analysis constitutes the numeration for the Minimalist component. This component now carries out a standard syntactic derivation and outputs logical and phonological representations of the input. The phonological representation is then compared against the original input string. If the match is good, processing is complete and the listener hears the two representations played simultaneously, otherwise the whole process starts again and a different candidate representation is generated.

5.6.2 Problems with the model

There are several problems with the model, arising perhaps, from the fact that Townsend and Bever have not yet worked out its mechanics in detail. One set of problems relates to the memory requirements of the model. It appears that temporary storage is required for two complete phonological representations of the same sentence. These representations must also presumably be stored in two separate buffers; otherwise, they would interfere with each other. However, the authors do not cite independent psychological evidence for (a) the requisite STM capacity; (b) the maintenance in STM of two distinct phonological representations of the same sentence or (c) the existence of two separate phonological buffers. Further, if these buffers have a limited capacity, as one would expect, it is hard to say how the model would cope with buffer overflow caused by excessive sentence length.

A second set of problems relates to the sequencing of processing events. When precisely does the listener hear the sentence? Only when a suitable match is found? What mechanism prevents the listener from hearing the sentence each time a comparison is made? And what if no match is found? Does the listener then not hear the sentence? From the information given in their book, Townsend and Bever also seem to predict a delay in hearing the sentence, spanning the time the input string initially enters the temporary store to the time a suitable match is found. The model also seems to predict variations in this delay, depending on the number of times candidate representations are generated before a satisfactory match is found. This would suggest that some sentences are heard systematically later than others relative to onset time. It is not clear if the authors wish to make that prediction.

There is a third set of problems concerning the memory requirements of the model in relation to the sequencing of processing events. If several approximations of the input are generated, a record must be kept concerning failed analyses. Otherwise the system runs the occasional risk of looping infinitely through the same wrong analyses. In the flow diagram on page 163 of Townsend and Bever's book, the authors present a box which indicates that data from the preceding analysis feeds into the preliminary analysis. It is not clear if this data contains a record of previous analyses and if so, what nature that record would take.

A key argument made by the authors is that there are an infinite number of possible sentences. This argument would seem to preclude the use of a strategy in which a token of some kind is stored in order to record each specific structure which has been proposed and rejected. Such a strategy depends on the forbidden assumption that it is possible to enumerate all possible structures. If sentence tokens are out of the question, does the record consist of entire sentence structures? If so, the storage requirements would be considerable. And if excessive on-line memory demands trigger an STM purge, records of previous analyses would be lost and the system would presumably get locked into a loop once again, repeating past errors indefinitely.

A fourth set of problems has to do with the fragmentary nature of conversational language in relation to the authors' claim that 'The sentence level is the fundamental object of language perception . . .' (2001: 5). Consider the following conversation heard recently on British radio:

Interviewer:	*are you on time*?
Interviewee:	*ish*
Interviewer:	*(laughs) are you on budget*?
Interviewee:	*ish*
Interviewer:	*(laughs)*

This example illustrates a difficult problem for the authors' model. The problem is that people communicate effortlessly without using complete sentences. It is not at all clear, given a sentence fragment as input, what the Connectionist component of the model would output. The Minimalist component seems to have two options. Given an incomplete numeration, it could 'crash'. It is not clear from the book what would happen to the parse then or what the behavioural correlates of 'crashing' might be. The other option is to accept the incomplete numeration, generate a complete tree structure and output a whole sentence. The difficulty is that it would not then be possible to phonologically match a complete sentence with a fragmentary input string.

And, in any case, it must be asked whether Minimalist principles are so subtle that they can convert the adjectival suffix -ish into an adverb meaning something like, 'approximately' and then, using material from the discourse context, build a tree structure, complete with IP node and all the rest, to derive sentences like 'I am approximately on time' and 'I am approximately on budget' and so on. It seems more plausible to regard the creative use of -ish here as the product of fluid verbal intelligence rather than something which a grammar can reasonably be expected to predict.

A fifth set of problems concerns the relationship between the Connectionist and Minimalist components. The entire argument of this seems to hinge on the ability of the Minimalist component to inform the Connectionist component in some way. However, it is not obvious from the flow diagram on page 163 of Townsend and Bever's book precisely how the Minimalist component informs the Connectionist component. There is a box which indicates that results from previous analyses feed into the preliminary analysis, but it is not made clear just what sort of information this box contains. The two components also use different representational formats: distributed representations versus symbolic representations. It is not clear how one format is translated into the other. In fact, there is something of a contradiction on this point, because, if translation between formats is possible, so that the Minimalist component can feed into the Connectionist component, this would appear to undermine the authors' argument (p. 147) that Connectionist models cannot represent detailed syntactic structure. If connectionist models can represent detailed syntactic structure, this would appear to negate the need for the Minimalist component.

5.6.3 Application to individual differences

Townsend and Bever's model is therefore currently unworkable. There is also a real problem in attempting to apply the model to individual differences. Chapter 3 showed that individuals differ in their ability to apply recursive rules. However, Townsend and Bever hardwire the ability to apply recursive rules by having a minimalist parser. It is therefore difficult for them to explain how individuals can differ in the ability to comprehend self-embedded sentences or why training should improve this ability.

5.7 Summary

The preceding discussion has indicated that current models of individual differences cannot fully explain the experimental data presented in Chapter 3. In the case of MacDonald and Christiansen's model, the problem is that SRNs do not fully implement rule-following behaviour. The ability to support rule-following behaviour is necessary in order to explain the ability to generalise to new structures as well as the ability to understand multiply-self-embedded sentences. In the case of Just and Carpenter and Caplan and Waters' models, the problem is that these models hardwire the ability to apply recursive rules into the language processing system. Consequently, it becomes difficult to explain why individuals need to be trained in order to apply such rules. The two hybrid models reviewed above also hardwire the ability to apply recursive rules and are therefore incapable of explaining individual differences in this respect.

The discussion thus far indicates that there are two basic issues which models of sentence comprehension have failed to resolve. Experience-based models have failed to incorporate STM processes, in particular, the conscious control processes that are necessary for rule-following behaviour. Rule-based models, on the other hand, tend to hardwire rule-following behaviour into the language processing system. Therefore, they cannot explain why language users need to be trained to follow rules. The inability of rule-based models to explain learning arises from the fact that these models fail to incorporate LTM processes. Hybrid models also fail to resolve these issues. The inability of each approach to explain STM or LTM processes reduces its ability to explain individual differences. A necessary first step in explaining individual differences therefore involves integrating LTM processes with conscious control processes. The next chapter reviews the literature on human memory and shows that this integration is not only possible, but has already been used by Ericsson and Kintsch (1995) as the basis for a theory of individual differences in cognitive skills. It will be argued that this theory, in combination with a compatible linguistic theory, can explain individual differences in the comprehension of complex sentences.

6
Long-Term Working Memory

6.1 Introduction

This chapter considers the relationship between sentence comprehension and the human memory system. It addresses two questions. First of all, does sentence comprehension involve long-term memory (LTM) traces of previously encountered sentences? This question is important because the experience-based approach treats sentence comprehension in terms of matching inputs with patterns stored in LTM. The rule-based approach, on the other hand, treats syntactic processing purely in terms of symbol manipulation in short-term memory (STM) or working memory. Secondly, does the limited capacity of STM or working memory constrain cognitive processes such as sentence comprehension? This question is important because the rule-based approach rests on a distinction between competence and performance. However, as argued by Lewis (1996), models of sentence comprehension seldom make reference to the psychology of memory. It is therefore important to find out if the assumption that working-memory capacity constrains comprehension really is justified by the psychological literature.

6.2 LTM for sentences

There was a debate in the 1970s to determine whether sentences that are heard or read are stored in LTM. The outcome of this debate had the following potential implications. According to the rule-based approach, sentence comprehension makes use of working memory as a temporary workspace. As soon as a sentence has been semantically interpreted, the surface form of the sentence is purged from memory. This idea would be supported by a failure to find evidence of LTM for surface form.

On the other hand, the experience-based approach assumes that language is learned from experience. If so, then one would expect all relevant aspects of sentences to be stored in LTM, including the surface forms of sentences. The attempt to determine the fate of sentence memory therefore touched on the central issues on which rule-based and experience-based approaches differ. Lack of evidence for long-term retention of surface form would support the rule-based approach and raise doubts about the experience-based approach. Conversely, evidence for the retention of surface form would support the experience-based approach and raise doubts about the rule-based approach. There were two distinct phases of this work. The first phase involved attempts to determine if there was long-term retention of the surface forms of sentences. The second phase involved attempts to determine the basis of syntactic priming.

6.2.1 The fate of surface structure

This review will first cover studies which support the view that surface structure information is purged from short-term memory (STM). These studies assumed the multi-store model of memory (Atkinson and Schiffrin, 1968), which is described later in the chapter. The review will then discuss studies showing that surface structure information is retained in LTM. These studies generally assumed a level of processing approach to memory (Craik and Lockhart, 1972), which is also described later in this chapter.

6.2.1.1 Evidence for an immediate purge of surface information

Sachs (1967) asked subjects to read passages and then decide whether a test sentence was the same or different from the one they had read in the passage. The test sentences were either the same as sentences appearing in the passage or differed from them in terms of form variables such as wording and grammatical voice and also in terms of meaning. For instance:

> **Original sentence**: There he met an archaeologist, John Carter, who urged him to join in the search for the tomb of King Tut.
> **Meaning change**: There he met an archaeologist, John Carter, and urged him to join in the search for the tomb of King Tut.
> **Form change**: There he met an archaeologist, John Carter, who urged that he join in the search for the tomb of King Tut.

Sachs varied the delay between presentation of the original sentence and presentation of the test sentence. With small delays, subjects could detect changes in both meaning and form. However, with longer delays, subjects could detect changes in meaning but not changes in form. This result was interpreted to show that verbatim information was only stored long enough for the meaning of a sentence to be computed. Thereafter, verbatim information was purged from STM. The meaning, on the other hand, was stored in LTM.

A similar result was obtained by Jarvella (1971). He asked subjects to listen to a recording of a passage which contained an identical sequence of words in either of two syntactic configurations. Each configuration consisted of three clauses. In the Short configuration, the *first* two clauses were structurally dependent. In the Long configuration, the *last* two clauses were structurally dependent. The different configurations are shown below with bracketing to indicate the pattern of dependency between the clauses:

> **Short configuration** [context (7 words) previous (6)] [immediate (7)]
> **Long configuration** [context (7 words)] [previous (6) immediate (7)]

An example of each configuration is shown below:

> **Short configuration**: [Kofach had been persuaded by the international to stack the meeting for McDonald.] [The union had even brought in outsiders.]
> **Long configuration**: [The confidence of Kofach was not unfounded.] [To stack the meeting for McDonald, the union had even brought in outsiders.]

Jarvella found that verbatim recall was high for both immediate and previous clauses in the Long configuration. However, recall was high only for the immediate clause in the Short configuration. In other words, recall was high for the second **sentence** in each condition regardless of its length. Jarvella concluded that the surface structure of a sentence is stored in STM only while the meaning of the sentence was being computed. Thereafter it was purged from memory.

6.2.1.2 Evidence of long-term retention of surface form

Several other findings raised questions about the results of Sachs and Jarvella. Moeser (1974) found that changes in surface form were detected better in sentences containing concrete rather than abstract words. Anderson (1974a) suggested that the recognition judgement task used by Sachs was not sufficiently sensitive to long-term storage of surface structure. Therefore, Anderson measured response latencies in a sentence verification task. In the task, subjects had to judge the truth or falsehood of a sentence which had been read earlier. He found that subjects were faster when verifying sentences whose surface form was unchanged compared to those whose surface form had been changed. He concluded that his findings argued against 'the short-term versus long-term memory distinction in favour of a model in which there are faster decay rates for the perceptual rather for than the propositional information.' (1974: 161). This is a reference to the levels of processing framework, which argued against separate STM and LTM stores in favour of a single system in which the recoverability of information depends on the manner in which that information has been encoded. Another implication of Anderson's study is that prior processing of a given sentence type facilitates its subsequent processing. This phenomenon will be discussed later in terms of syntactic priming.

Graesser and Mandler (1975) asked subjects to carry out two different sentence-processing tasks. Subjects either carried out a comprehension task or a grammaticality judgement task. In a forced choice recognition test, verbatim memory was found to be poor for the semantic task but high for the grammaticality judgement task. It was concluded that the level of processing (that is, semantic processing versus processing of surface form) determined the memorability of surface structure. McDaniel (1981) varied syntactic complexity and found better surface structure memory for more complex sentences. An interesting aspect of McDaniel's study was that the complex sentences which he used were self-embedded sentences. His study therefore raises questions about the idea that the processing of self-embedded sentences is hindered by limitations in STM.

Kintsch and Bates (1977) found that surface memory for classroom lectures can be retained for two days, but is reduced after three days. They also found that verbatim memory for extraneous comments made during a lecture was higher than for other types of statements. According to Fletcher (1994), this result may have been due to the distinctiveness of extraneous comments, such as jokes and announcements, relative to the less distinctive contents of the lecture itself. Keenan *et al.* (1977) made a distinction between statements of high versus low interactional content, for example:

> **Low interactional content**: I think there are two fundamental tasks in this study.
> **High interactional content**: I think you've made a fundamental error in this study.

It was found that recognition memory was higher for high interactional content statements compared with low interactional content statements. Bates *et al.* (1978) also found that recognition memory for surface structure was higher for explicit references than for anaphoric utterances. This result was replicated by Stevenson (1988). Stevenson also noted that meaning-preserving changes in surface form in the Bates *et al.* study appeared to interfere with discourse expectations in ways which improved surface memory.

In addition to the experimental studies reviewed above, there are also logical considerations which indicate that the surface forms of sentences have to be retained in LTM. Firstly, language users can recognise voices, in the case of spoken languages and handwriting and fonts in the case of written language. It is not clear how this recognition can occur in the absence of memory traces for surface forms. Secondly, language consists of numerous fixed expressions, such as multi-word units, clichés, figures of speech, idioms, quotations and so on, which have invariable surface forms. Again, it is not immediately clear how such expressions could be recognised as such if language users did not store surface forms in LTM. It is possible to suggest a 'dual route' type theory of memory, whereby the surface forms of free expressions are purged from STM while those of fixed expressions are stored in LTM. However, this raises the question of how the language processing system can tell in advance whether an expression is free or fixed and should therefore be purged from STM or stored in LTM.

There were many more studies carried out on this topic than have been reviewed here. However, the studies reviewed here are representative. Essentially, long-term retention of surface structure information can be detected depending on the types of materials used and the kinds of processing which are carried out. The research on LTM for surface form was therefore supportive of the experience-based approach. As indicated earlier, some of these studies showed that previous sentence processing episodes could facilitate future sentence processing. This last issue has been investigated more recently in the area of syntactic priming. Findings in this area are reviewed next.

6.2.2 Syntactic priming

Syntactic priming is a phenomenon whereby exposure to a given syntactic structure facilitates the subsequent comprehension or production of that structure. It has been suggested that syntactic priming provides evidence for a purely syntactic level of representation (Bock and Loebell, 1990; Branigan *et al.*, 1995; Frazier, 1995). The research on syntactic priming has been concerned with establishing whether there are, in fact, purely syntactic priming effects which can clearly be distinguished from effects of non-syntactic priming. If this is found to be the case, then such evidence might constitute support for a rule-based approach to sentence processing. However, as noted in Chapter 3, evidence of a level of pure syntactic representations does not distinguish between rule-generated syntactic structures and rote-learned syntactic forms. The following review will (a) assess the evidence for purely syntactic priming and (b) evaluate the extent to which such evidence makes it possible to discriminate between rule-generated and rote-learned syntactic structures.

6.2.2.1 Priming at different levels of representation

Frazier *et al.* (1984) found that the second conjunct of a conjoined structure was read faster than the first if both conjuncts shared certain structural features. The priming effect was found when (a) both conjuncts had the same syntactic structure; or (b) the same thematic structure or (c) when the animacy feature of nouns in both conjuncts was the same. This study therefore found priming at both syntactic and non-syntactic levels. Additionally, it was difficult to say whether the priming was purely syntactic and not due to thematic or other sources of priming, such as lexical similarity and rhythm.

One finding from the study is of particular interest. It was found that the priming effect was greater for marked structures, such as shifted heavy NP structures and non-minimal attachment structures. This suggests that relatively unfamiliar structures have a stronger priming effect than relatively familiar structures. The authors suggested that priming might result from the parser making use of representations already constructed. In the experiment, the target clause followed immediately after the prime clause (since prime and target were conjoined clauses), in which case it might be that the representation constructed for the prime was still available in STM and could therefore be reused to process the target. However, in cases where an unrelated sentence separates the prime and target, the representation of the prime sentence ought to have been purged from working memory. If priming still occurs in that case, then priming might be a result of a representation of the target which has been stored in LTM. Such a finding would be more consistent with the experience-based than the rule-based approach. Evidence for long-term syntactic priming will be presented later.

6.2.2.2 Evidence for pure syntactic priming in production

Bock and Loebell (1990) used the following kinds of sentences as primes:

> **Prepositional dative**: The wealthy widow gave an old Mercedes to the church.
> **Prepositional locative**: The wealthy widow drove an old Mercedes to the church.
> **Double object dative**: The wealthy widow sold the church an old Mercedes.

The prepositional dative and prepositional locative have the same syntactic structure but different thematic structures. The double object dative has a different syntactic structure compared with the other two structures but it has the same thematic structure as the prepositional dative. Subjects had to repeat one of the three sentence types, ostensibly in preparation for a memory task. After repeating one of the three sentence types, subjects had to describe in words a picture depicting a dative event. Distracter sentences intervened between the prime sentence and the picture. Subjects produced equal numbers of prepositional dative responses whether the prime was a prepositional dative or a prepositional locative but they produced fewer prepositional datives following double object primes. The result was interpreted to show a pure syntactic priming effect, since, 'Conceptual similarity was no more likely than conceptual dissimilarity to lead to structural repetition' (16).

However, the thematic structures associated with locative and dative verbs are similar in that both types of relation involve a change of location. So it is not clear from the results of this experiment that priming is purely structural. Bock and Loebell accepted this possibility and carried out another study where they hoped to cancel the conceptual similarities between the different structures they used. This second experiment used the following structures:

> **Passive**: The 747 was alerted by the control tower.
> **Locative**: The 747 was landing by the control tower.

Using these materials, results similar to those in the first experiment were found. That is, passive and locative primes were equally likely to induce subjects to produce passive responses. However, it is still not certain that Bock and Loebell effectively cancelled the conceptual similarities between passives and locatives. Firstly, some of their materials are ambiguous between locative and passive readings. For instance:

> **Passive**: The construction worker was hit by the bulldozer.
> **Locative**: The construction worker was digging by the bulldozer.

The passive sentence has both a passive and locative reading; that is, a bulldozer may have hit the worker, or the worker may have been hit by someone near a bulldozer. Secondly, all the materials are temporarily ambiguous between passive and locative readings. In the sentences above, the reader is unsure, up to the time he or she reads the auxiliary verb, whether the sentence is passive or locative. In the passive sentences, the ambiguity persists up to 'by'. If both readings are activated during the comprehension of the prime sentences, there is a possibility that both readings might exert an influence on the production of the target sentence.

It is therefore not certain from Bock and Loebell's study whether or not they obtained evidence of pure syntactic priming. A follow-up study by Heydel and Murray (1997) led to the conclusion that, 'Conceptual form and syntactic structure appear to interact, with conceptual form always playing an important part, while the role played by syntax may vary with the task'(6). If it is assumed that the Bock and Loebell study produced a purely syntactic priming effect, the question remains whether this priming was due to rule-generated or schematic syntactic structures. Bock and Loebell acknowledge that their study does not indicate whether priming occurs at the level of the sentence or of substructures generated by individual syntactic rules:

> It is unclear at what level of the phrase structure configuration the priming effects emerge. Though we have focused on sentential configurations, we doubt that priming is restricted in this way. It should arise at all levels (33).

As yet, there is no evidence to clarify the issue.

6.2.2.3 Evidence for pure syntactic priming in comprehension

Pickering and Branigan (1998) found that when the same verb was used in both prime and target, subjects produced 17.2 per cent more target completions that were of the same type as the prime than target completions that were of the alternative type to the prime. This percentage dropped to 4.4 per cent when the sentences contained different verbs. Priming with different verbs was stronger only when two priming sentences were used before each target sentence. However, the study did not control for thematic structure, so it is not certain that the priming effect was purely syntactic. Since (a) thematic structure might be encoded in the lexical entry of a verb and (b) a stronger priming effect was found when the verb was repeated, it is possible that the effect found was largely thematic. The authors also acknowledge that their results do not distinguish between syntactic rules and subcategorisation frames. Given that subcategorisation frames are fixed schemata, the results are, at best, ambiguous between rule-generated and schematic syntactic structures.

6.2.2.4 Long-term syntactic priming

Cuetos *et al.* (1996) carried out an experiment which provided evidence for long-term syntactic priming. Two groups of Spanish seven-year-olds were asked to read, over a two-week period, stories containing sentences which were biased towards a high attachment or low attachment relative clause interpretation. For instance, in the sentence 'The daughter of the colonel with the limp', the relative clause 'with the limp' can be attached to the noun phrase headed by 'the daughter' (high attachment) or to 'the colonel' (low attachment). After the two weeks elapsed, the children received no materials concerned with the research for a week, after which they were tested for attachment preference. It was found that the children who had been exposed to high attachment bias materials tended to prefer high over low attachment relative to their pre-test performance. No effect of the intervention was found for the children given the low attachment biased materials however. This result was explained by saying that, while the research was under way, the children were exposed to materials with a high attachment bias (a high attachment preference in Spanish had been established in an earlier study). If priming is a long-term effect, then it cannot be explained in terms of the temporary raising of activation in a procedure. It seems more plausible to suggest that exposure to a certain structure leads to a long-term change in the language processing system in a manner akin to learning. The last study to be reviewed in this section suggests that syntactic priming is a form of learning.

6.2.2.5 A connectionist model of long-term syntactic priming

Cuetos *et al.*'s findings were interpreted as providing support for the authors' tuning hypothesis. According to this hypothesis, parsing preferences reflect the statistical regularities found in natural language. If tuning can be equated with priming, then priming can be described as a kind of statistically driven learning, as suggested by Bock *et al.* (1996).

Bock *et al.* (1996: 8) found an effect of priming even when ten unrelated sentences intervened between prime and target. It was concluded that:

> The persistence of structural priming over 10 unrelated sentences drives home the point that these effects are not transient, and cannot be attributed to a momentary change of activation. We need to consider an alternative mechanism for the priming effect, one that entails a more persistent change in the processing system.

The authors propose the following account of priming:

> [...] Structural priming can be seen as a dynamic vestige of the process of learning to perform language. We call this process "learning to talk", in a very literal sense. It is not learning language, but learning to produce it. So "learning to talk" is learning procedures for efficiently formulating and producing utterances. What structural priming shows is that these procedures undergo fine-tuning in every episode of language production (11).

6.2.3 Summary

Taken together, the work on LTM for sentences shows that previous sentence processing episodes can influence subsequent sentence processing. It therefore appears that sentence processing makes use of long-term sentence memory. This finding is theoretically relevant in two ways. Firstly, proponents of the rule-based approach have criticised the notion that knowledge of language includes memory for previously encountered sentences. Evidence for long-term retention of sentences weakens this argument, particularly in view of the fact that previous sentence processing facilitates subsequent processing. Secondly, according to the experience-based view, knowledge of language is learned from experience. This argument would be less plausible if there was no LTM for sentences. Long-term retention of sentences therefore supports the argument that knowledge of language is learned.

6.3 LTM contribution to STM/working memory

This section considers the assumption that limitations in the capacity of STM or working memory constrain cognitive processes, such as sentence comprehension. This assumption underlies the competence–performance distinction which is crucial to the rule-based approach.

6.3.1 The channel capacity model of human memory

Chapter 2 described how information theory formed the basis for Hockett's finite state model. Information theory also formed the basis for theories about human memory. In particular, it was thought that the notion of channel capacity could also be applied to human beings. That notion was used by Shannon to denote the capacity of a physical medium, such as electric cable, to transmit signals. Psychologists applied the idea to human psychology as follows:

> A secretary, for example, taking shorthand and later converting it into a typed letter, can be viewed as a communication system. The boss' original speech is the "source" of the "message", which undergoes two encoding conversions if the secretary is viewed as a communication channel. The first is the transcribed shorthand, the second is typing with the typed copy as the "output". Whoever reads the typed copy may be viewed as the "destination". In both conversions, it may be meaningful to talk of "channel capacity" as the upper rate limits or top speed that the secretary can perform errorless transcription and errorless typed output (Lachman *et al.*, 1979: 65).

This application of the notion of 'channel' is highly context-dependent. A secretary is a channel insofar as he or she is transcribing information. If the secretary were simply reading a typed memo, then he or she would not be functioning as a channel. It would seem better to have a non-relative application of the idea. The application of 'channel capacity' to humans therefore seems a rather unsatisfactory use of information theory, and the discussion will indicate that there was another, more sophisticated and possibly more productive, application of the theory. Nevertheless, the simplistic application described above was taken seriously and used as the basis for experimental work. Early research did, in fact, suggest that human beings behaved as limited capacity channels. This research concerned (a) choice reaction time; (b) language processing; (c) the span of absolute judgements; and (d) the span of human immediate memory.

6.3.1.1 Choice reaction time

One of the earliest applications of information theory in psychology was in the area of choice reaction time. The experimental paradigm used in this area involved getting subjects to respond to one of several possible events by pressing a key. The number of alternative events was varied and the time taken to respond, given a specific number of alternatives (thus 'choice reaction time'), was measured. In terms of information theory, the greater the number of alternatives, the greater the amount of information transmitted by the subject. Using himself as a subject, Hick (1952) found that choice reaction time increased with an increasing number of alternatives. Hick concluded that, 'Fairly strong evidence has been obtained that the amount of information extracted is proportional to the amount of time taken to extract it . . .' (25). This finding was replicated by Hyman (1953).

However, Fitts and Switzer (1962) found that, while choice reaction time was a function of set size, as indicated by Hick (1952) and Hyman (1953), it was also a function of set familiarity. Subjects were found to respond faster to stimuli from known compared to unknown sets. Fitts and Switzer also found that reaction time was a function of inferred rather than actual set size. In other words, if subjects assumed (incorrectly) that a set was larger than it actually was, their responses were slower. These findings were taken to mean that channel capacity was not fixed but reflected human adaptivity. Mowbray and Rhoades (1959) also found that choice reaction time decreases with practice. They gave one subject reaction time tasks involving two or four alternatives. Initially, the subject's reactions were faster in the two-alternatives condition than in the four-alternatives condition as expected from Hick's (1952) study. However, reaction times for both tasks decreased until they become equal after 13 trials. Subsequently, reaction times for both tasks decreased asymptotically.

6.3.1.2 The span of absolute judgements

Research on absolute judgements initially promised to provide quantitative measurements of human channel capacity. In these experiments, subjects had to identify stimuli which varied along single or multiple dimensions. For instance, a subject might be presented with several tones and asked to respond to each individual tone with a unique numeral. When the number of alternative tones per set was small, it was easy to categorise each tone correctly. However, when the number of alternative tones was larger, subjects began to make errors of categorisation, or 'judgement'. The maximum number of different tones from which subjects could make accurate judgements therefore provided a measure of channel capacity in bits. Miller (1956) reviewed studies which showed that, for uni-dimensional stimuli such as tones and tastes (for example, saltiness), subjects could discriminate accurately between 6.5 categories of stimulus intensity. Miller therefore calculated that humans had a channel capacity of 2.6 bits.

However, Miller reported two surprising findings. Firstly, he mentioned evidence that musically sophisticated individuals can discriminate between 50 and 60 different pitches. He commented, 'Fortunately, I do not have the time to discuss these remarkable exceptions. I say it is fortunate because I do not know how to explain their superior performance' (1956: 84). Miller also reports that increasing the dimensionality of a stimulus leads to an *increase* in channel capacity. For instance, it was found that for six-dimensional acoustic stimuli, the channel capacity was 7.2 bits. The fact that channel capacity increased with (a) domain expertise and (b) the richness of the stimulus was clearly problematic for the idea that channel capacity is fixed.

6.3.1.3 The span of immediate memory

Another application of the notion of channel capacity was in the area of immediate memory (short-term recall). Miller reported that humans have an average immediate memory span of around seven decimal digits. Since each digit has an information value of 3.3 bits (because each digit belongs to a set of ten alternative digits, that is, 0 to 9), it follows that the span of immediate memory should be about 23 bits (7 items multiplied by 3.3 bits). Miller then stated, on the basis of previous studies, that isolated English words have an information value of 10 bits per word. If the maximum amount of information which can be retained is 23 bits, it follows that the span of immediate memory for isolated English words is 2.3 words (23 bits divided by 10 bits).

However, Miller reported two experiments which showed that the span of immediate memory for words and a variety of other materials is 7 rather than 2.3 words. This means that the span of immediate memory is not a function of the amount of information in the stimulus. Rather, the size of immediate memory is given in terms of a fixed number of items (or chunks), regardless of the size of each item. Miller (1956: 93) explained the unexpected result in terms of an increase in the efficiency of the channel through recoding:

> We are dealing here with a process of organising or grouping the input into familiar units or chunks, and a great deal of learning has gone into the formation of these familiar units [...] Since the memory span is a fixed number of chunks, we can increase the number of bits of information that it contains simply by building larger and larger chunks, each chunk containing more information than before.

Miller illustrated this recoding process in terms of a Morse code operator. The operator may begin with chunks which correspond to individual signals of the Morse code. In time, the operator learns to treat signals as letters, then letters as words, and finally, he can treat the words as phrases. Thus the operator 'learns to increase the bits per chunk' (1956: 93). Miller then goes on to report experiments which show that the number of items in immediate memory can be increased by grouping items and applying a name to the group.

This process is similar to the way in which the code of an information system can treat recurrent sequences of messages as single units which can be encoded using one signal ('block coding'). Miller suggests that humans can recode material in a hierarchical fashion, by grouping the names of each group into a super-group and then applying a name to the super-group and so on. Through intensive practice, subjects can dramatically increase memory span using hierarchical encoding schemes. Miller reports that a fellow researcher learned to increase his memory span to 40 digits by using a hierarchical encoding scheme. In sum, therefore, the research reported in Miller (1956: 95) shows that capacity limitations can be overcome:

> [...] The span of absolute judgement and the span of immediate memory impose severe limitations on the amount of information that we are able to perceive, process and remember. By organising the stimulus input simultaneously into several dimensions and successively into a sequence of chunks, we manage to break (or at least stretch) this informational bottleneck.

In this passage, Miller's comments are suggestive of an alternative and possibly more productive application of information theory. An important concern for information theory was to find ways of making communications systems more efficient. For instance, information channels, such as electrical cable, introduce interference into transmitted signals which can cause one signal to be mistaken for another. According to Pierce (1980: 44), the problem which Shannon set himself was 'how to represent or encode messages from the message source so as to attain the fastest possible transmission over the noisy channel [...] without error'. Apparently, the solution to this problem involves designing efficient codes (Pierce, 1980: 276). This aspect of information theory suggests the following analogue.

It would be easy to suppose that the speed of communication through a channel is a function of the physical capacity of the channel.

However, this would be a mistake, given that, to a large degree, the speed of information depends on the efficiency of the ways in which messages are encoded. It may be that psychologists studying human memory were making the same sort of error, that is, mistaking the capacity of the coding system for the capacity of the brain. In other words, measurements of STM capacity might be better thought of as measuring the task-specific efficiency of the representations used by subjects, rather than measuring their absolute mental capacity.

Subsequent research did, in fact, confirm the idea that immediate memory does not have a fixed capacity. For instance, Mackworth (1963) found that memory span is affected by the nature of the materials used, while Conrad and Hull (1964) found that memory span is related to the probability of acoustic confusion. A potentially productive line of psychological research would therefore have been to explore the efficiency of different kinds of codes for various mental tasks. It could, in fact, be argued that learning involves the development of progressively more efficient codes for representing input from the relevant domain. For some reason, this promising line of research, which seems a much better application of information theory to psychology, was not taken up. Instead, new formulations of the notion of capacity limitations have been proposed regularly from the 1950s to the present. As the discussion below shows, each of these new formulations was always subsequently undermined by evidence that so-called capacity limitations are related to learning and to the manner in which information is encoded in LTM.

6.3.2 The multi-stage model of human memory

According to Lachman *et al.* (1979), there was a shift in the 1960s from viewing human beings as communication systems towards viewing them as computational systems. In memory theory, this shift was marked by Atkinson and Schiffrin's (1968) model of human memory that was based on the computer metaphor. According to the multi-store model, human memory consists of three main stores: various sensory stores, the short-term store and the long-term store. These stores correspond, respectively, to the following computer components: input buffers, random access memory and disk storage.

The sensory stores are specific to the sensory modalities. Information which enters the sensory stores through the senses is held very briefly. Some of it is attended to and processed within the short-term store. Some of the information which is processed in the short-term store is then transferred to the long-term store (Eysenck and Keane, 1995). Only the short-term and long-term stores are relevant for this discussion; therefore, details about the sensory stores will be passed over.

6.3.2.1 Differences between STM and LTM

Traditionally, STM and LTM are thought to differ in several respects. STM has a small capacity, about seven chunks (Miller, 1956) and information is quickly lost when attention is diverted (Peterson and Peterson, 1959). The short-term store is also associated with conscious processes for controlling the flow of information. These processes include rehearsal, naming (that is, attaching a verbal label to a stimulus), grouping (for example, perceiving a group of letters as a single word) and recoding (that is, concatenating the material to be remembered according to pronounceability or meaningfulness). Retrieval from STM is thought to be automatic and rapid. LTM, on the other hand, has a vast capacity and it stores information durably in the form of an associative network (Atkinson and Schiffrin, 1968; Waugh and Norman, 1965). Encoding information into LTM is said to be slow and probabilistic (Simon, 1973). Retrieval from LTM is also thought to be slow and inefficient because of the scarcity of retrieval cues.

6.3.2.2 STM and cognition

Ericsson and Kintsch (1995) propose that the hypothesised differences in the accessibility of information from STM and LTM have led researchers to suppose that cognitive processing is based on information in the short-term store. This is because retrieval from the short-term store is thought to be rapid, automatic and accurate, while access to and from the long-term store is slow and inaccurate. The distinction between STM and LTM is an important one for the rule-based approach. It underwrites the assumption that the process of sentence comprehension is constrained by limitations in STM. For instance, it is thought that STM limitations impose strict restrictions on the application of recursive rules. The limited capacity of the short-term store therefore underlies the competence–performance distinction which enables proponents of the rule-based approach to explain the manifest inability of language users to comprehend sentences beyond a certain level of structural complexity.

However, there are serious problems with the notion of STM capacity limitations. These problems have the potential to invalidate the competence–performance distinction. Ericsson and Kintsch (1995) have presented some empirical arguments which show that, during skilled processing, access to LTM can be fast and accurate. Given that LTM has a vast capacity, the notion of capacity limitations consequently becomes very insecure. Ericsson and Kintsch use these arguments to propose a re-thinking of the entire relationship between STM and LTM. For the moment, the discussion will focus on these arguments and then turn towards Ericsson and Kintsch's alternative formulation towards the end chapter. The next four paragraphs are based on the argument presented in Ericsson and Kintsch (1995).

6.3.2.3 LTM access during skilled processing

Firstly, Ericsson and Kintsch note that information can be stored rapidly and accurately in LTM during skilled processing. For instance, Craik and Lockhart (1972) and Hyde and Jenkins (1973) found that intentional and incidental LTM storage during skilled processing led to equivalent levels of recall. de Groot (1978) found that chess masters who had just made an unfamiliar chess move could reproduce almost the entire chess position in verbal reports of their thought processes during the selection of the move. Lane and Robertson (1979) also found that chess experts' recall in experiments such as de Groot's was just as good as when they had been informed of the memory test in advance. Charness (1979) and Engle and Bukstel (1978/1977) obtained the same findings for bridge, with incidental recall increasing with the level of expertise. Norman *et al.* (1989) found that incidental memory after a diagnosis was greater for medical experts than for medical students. According to Ericsson and Kintsch, the findings presented above indicate that a lot of relevant information is stored in LTM as an integral part of carrying out a skilled activity. This information can be retrieved quickly and accurately because there are more cues for retrieval from LTM in skilled processing compared with domain-neutral laboratory tasks.

6.3.2.4 Fast and reliable retrieval of familiar material from LTM

Secondly, Ericsson and Kintsch question the view that selective retrieval from STM is necessarily faster and more accurate than retrieval from LTM. Instead, it appears that selective retrieval from LTM is rapid and accurate if the material to be retrieved is familiar and overlearned. For instance, selective retrieval from STM is thought to take about one second while LTM retrieval time for unrelated paired associates (foreign language vocabulary items) is two seconds (Crutcher, 1990). After extensive practice, LTM retrieval time is reduced to one second (Crutcher, 1992). For related paired associates (for example, monosyllabic English words) retrieval times are 800 ms after the first study trial and 400 ms after 18 practice blocks. Thus, retrieval time from LTM for highly familiar and overlearned information can be twice as fast as retrieval from STM for unfamiliar materials.

6.3.2.5 Efficient LTM storage

Thirdly, Ericsson and Kintsch question the view that increases in recall during skilled activities arise entirely due to chunking in STM. They cite Chase and Simon (1973), who found that experts had superior recall for chess materials than chess novices. This superiority was initially explained in terms of chunking. However, Charness (1976) found that an interpolated task did not affect chess memory. This result was taken to show that chess experts were storing information in LTM, given that information in STM is rapidly lost if unattended. Ericsson and Kintsch also report experiments of their own which show that interpolated activities do not lead to a loss of text memory.

Therefore, it appears that (a) information is stored rapidly and accurately in LTM during skilled processing, (b) information can be recalled rapidly and accurately from LTM if that information is highly familiar and (c) the ability of experts to store and retrieve large amounts of information is not due to chunking in STM but to efficient storage and retrieval from LTM. In addition to the arguments presented by Ericsson and Kintsch, other researchers have made observations which lead to serious questions about the construct of STM.

6.3.2.6 Problems with the concept of STM

For instance, Wickelgren (1965: 53) obtained evidence to suggest that 'STM and LTM are performed by the *same* system operating in a qualitatively different manner under different degrees of learning'. Subjects heard a list of four letters, followed immediately by a list of eight letters. The task was to write down the eight letters and then recall the four original letters. Wickelgren found that recall of the four letters was influenced by the degree of similarity between the two lists. Recall was lower when the two lists were similar compared with when they were different. This result indicated that the second list interfered with the first. The interference could be explained by saying that subjects were learning both lists incidentally in LTM. A study by Bower and Winzenz (1969) confirmed this interpretation. Bower and Winzenz presented subjects with a series of digit strings for immediate serial recall. They repeated certain strings surreptitiously several times and found that short-term recall increased gradually for the repeated strings compared with the non-repeated strings. This finding indicates, once again, that STM tasks involve LTM storage.

Chi (1976) reviewed the literature on STM to find out if STM capacity increases with age. She found no evidence to conclude that STM capacity varied with age. However, she found that children did not have the kinds of conscious processing strategies used by adults. The strategies she studied were rehearsal, naming, grouping and recoding. Chi also found that children and adults differed in the contents of LTM in ways which allowed adults to display superior recall. Chi studied three kinds of LTM differences and concluded the following. Firstly, the chunks possessed by adults are generally larger than those possessed by children. For instance, an adult can recognise a word as a single chunk whereas a child might perceive a word as a series of letters. Secondly, adults have more recognisable chunks than children. Thirdly, adults may possess richer associations between chunks, making it easier for them to access specific chunks. Chi reports findings that, where children and adults possess similar chunks, the richer associative links possessed by adults enable them to access their chunks faster than children.

The multi-stage model was also criticised by Craik and Lockhart (1972), who proposed that it is the level at which information is processed which determines how well it can be remembered, rather than on the size or accessibility of the memory stores. For instance, attending to the semantic attributes of a word was found to be more effective for later recall than attending to its perceptual attributes (for example, counting the number of letters in the word). The explanation of this result was that semantic encoding represents a deeper level of processing than perceptual processing. Thus, it was argued that the distinction between STM and LTM was based on the depth of processing rather than on the existence of different memory stores.

However, Craik and Tulving (1975) found that depth of processing was not the only determinant of long-term retention. Information which was processed in a more elaborate fashion was found to be recalled better than information which was processed in a less elaborate manner. Eysenck (1979) also argued that distinctiveness was an important factor: memory traces which were unique in some way were found to be easier to recall than memory traces which were similar to other traces. What all these findings seemed to indicate was that the way in which information is encoded determines the way in which it is recalled. This observation was made by Wiseman and Tulving (1976) who referred to it as the encoding specificity theory.

6.3.3 The phonological loop memory model

As described in the previous chapter, Baddeley and Hitch (1974) replaced the notion of STM with that of working memory. The reason for this replacement was that STM refers to the storage of information and does not include the manipulation of information. The concept of working memory was meant to incorporate both the temporary storage and manipulation of information. Working memory as envisaged by Baddeley consists of three components: a phonological loop, a visuo-spatial sketch pad and a central executive. The phonological loop and sketch pad perform storage functions while the central executive performs processing functions.

Baddeley *et al.* (1975) proposed that the capacity of the phonological loop is determined by the time taken for phonological representations in STM to decay. They obtained evidence which suggested that verbal STM is inversely related to word length, a finding also reported by Cowan *et al.* (1992). Baddeley *et al.* explained the word-length effect by proposing that phonological representations are stored in a 'phonological loop' – an output buffer for speech production – which comprises part of the more complex working-memory system. This 'phonological loop' theory suggested that there may be clear-cut limits to the capacity of STM after all.

However, this idea has been challenged by Service (1998), who found that memory span is affected by phonological complexity (that is, the number of phonemes) rather than just word length. More definitive evidence against the validity of the phonological loop theory has recently been reported by Lovatt *et al.* (2000), who found that the word-length effect was not obtained when factors such as number of phonemes, phonological similarity and frequency were controlled for. Further evidence against the phonological loop theory is reported by Hulme *et al.* (1991), who found that words are recalled better than non-words. Additionally, Roodenrys *et al.* (1994) found that more frequent words are better recalled than less frequent words. The effects of word familiarity and frequency point to a LTM contribution to STM.

6.3.4 Summary

In summary, the strict separation of STM and LTM has been consistently undermined by findings that LTM contributes to STM. These findings also undermine the idea that cognitive processes are constrained by the limited capacity of STM or working memory. It appears that the capacity of working memory can be boosted by making efficient use of LTM. These findings have negative implications for the competence–performance distinction, which rests on the idea that language comprehension is constrained by the limited capacity of STM. The possibility arises that putative effects of STM or working memory on sentence comprehension are, in fact, effects of insufficient learning, leading to inadequate encoding of linguistic information. It is therefore possible that individual differences in comprehension are due to individual differences in procedural syntactic competence. This argument has been developed by Ericsson and Kintsch (1995), whose theory is described below.

Ericsson and Kintsch's theory lies outside mainstream sentence processing research. It will be argued that the theory addresses the key issues which cause problems for experience-based and rule-based theories and combines the positive elements of each. The theory incorporates the experience-based view that knowledge of language is learned from experience and stored in LTM as an associative network. The theory also incorporates the rule-based view that human beings are capable of constructing complex structured representations. But whereas the experience-based view focuses on LTM processing and the rule-based view focuses on STM processes, Ericsson and Kintsch's theory treats cognitive processing as an activity which involves the co-ordination of STM and LTM processes.

6.4 Ericsson and Kintsch's theory of skilled processing

Ericsson and Kintsch (1995) agree with Just and Carpenter (1992) that there exists a limited capacity working memory, which they call short-term working memory (ST-WM). However, they argue that it is possible to extend the capacity of working memory by making efficient use of LTM. The extended working memory is referred to as long-term working memory (LT-WM). The difference between ST-WM and LT-WM is that storage in ST-WM is temporary but it provides immediate access whereas storage in LT-WM is durable but it requires stable cues for reliable access. It is possible to create stable systems of retrieval cues, called retrieval structures via control processes which operate on long-term associative memory in order to construct efficient representations that cater for the informational requirements of given cognitive tasks.

6.4.1 Retrieval structures

Ericsson and Kintsch's theory is partly based on an earlier theory by Chase and Ericsson (1982) which sought to explain a 1000 per cent increase in digit span displayed by subjects who had undergone extensive training in mnemonic techniques. It appeared that subjects were able to increase their digit span by recoding the material to be remembered in terms of a hierarchical system of cues, such as the one described by Miller (1956). The use of such retrieval structures apparently enabled subjects to subsequently recall the digits in any order required by the experimenter. This highly flexible behaviour is difficult to account for in connectionist terms because it is not simply due to associative learning. Rather it is the result of deliberately creating and manipulating associative links in accordance with a plan of action. In some ways, retrieval structures are similar to data structures used by computer programs in order to represent information in a task-specific manner.

The enhanced availability of information due to the use of retrieval structures is not confined to recall tasks. Ericsson and Kintsch argue that experts in various domains of skill develop task-specific retrieval structures in order to facilitate their performance. They review studies which show that the superior working-memory capacity displayed by experts is due to the use of retrieval structures. These findings have been obtained in domains such as mental multiplication, memory for dinner orders, medical expertise and chess. The studies in chess are particularly significant.

There are, apparently, chess masters who have trained themselves to play blindfold chess. In this type of game, the chess master is blindfolded and depends on verbal communication to find out about the opponent's move and to make his or her own moves. The chess master therefore has to mentally represent the continually changing chess board throughout the game. Ericsson and Kintsch argue that this feat is made possible by the use of a retrieval structure corresponding to a mental chess board. What is even more astonishing, however, is that some chess masters can play ten games of blindfold chess simultaneously and there is even one chess master who can play 30 games of blindfold chess simultaneously. Reports of such feats are often accompanied by assurances that these high levels of skill are due to training rather than innate talent for chess.

6.4.2 Retrieval structures as systems of representation

Retrieval structures appear to be a task-specific representation of a domain. That is to say, they are not merely mnemonic techniques for retrieving large amounts of information. Rather, they capture the structure of the task in question and Ericsson and Kintsch use the term 'knowledge structure' to denote this function. The creation of knowledge structures depends on thorough familiarity with the task in order to anticipate its memory demands. The development of knowledge structures also involves both learning from experience and deliberate control of cognitive processes. In other words, it involves both LTM, which is the focus of experience-based models and the control processes that underlie rule-governed behaviour. Given that the storage capacity of LTM is vast and the creative potential of rule-governed processes is limitless, Ericsson and Kintsch's theory suggests that the human cognitive capacity is vast.

6.4.3 Individual differences in linguistic representation

Ericsson and Kintsch therefore disagree with Just and Carpenter on the proposal that individual differences in comprehension are due to inherent differences in the size of ST-WM. For Ericsson and Kintsch, the major source of individual differences in comprehension is in the ability to encode linguistic material efficiently in LTM. This ability allows more skilled individuals to use LTM as an extension of their working memory. The best evidence for their approach, according to Ericsson and Kintsch, comes from studies which compare reading ability with domain knowledge.

In such studies, it has been found that individuals with good reading skills outperform individuals with poor reading skills when texts do not require domain knowledge. It has been suggested that good reading skills depend on having a large working memory. However, when texts require domain knowledge (such as knowledge of football, for instance), individuals with poor reading ability but high domain knowledge outperform individuals with high reading ability and poor domain knowledge. Differences in the ability to comprehend texts are therefore not dependent on inherent differences in working-memory capacity. Rather these differences arise from differences in the ability to encode information in LTM efficiently. Football experts have presumably developed mental models of football which enable them to comprehend texts about football better than football novices. On the other hand, skilled readers may have developed mental models for texts and these models may enable them to comprehend domain-neutral texts better than less skilled readers.

The following paragraphs consider some implications of Ericsson and Kintsch's theory for the models proposed by Just and Carpenter (1992), Caplan and Waters (1999) and MacDonald and Christiansen (2002).

6.4.4 Implications for Just and Carpenter's model

It appears that individual differences in working-memory capacity as measured by Daneman and Carpenter's test are related to individual differences in the ability to make efficient use of LTM. Daneman and Carpenter (1980) found evidence that subjects in the reading span test actively tried to encode associations between last words of sentences and to reconstruct sentences from gist. Ericsson and Kintsch interpret such efforts as evidence that performance in the reading span test depends on the ability to encode information in LTM in ways which facilitate retrieval. Supporting evidence for this interpretation is the fact that reading span is correlated with recognition memory (Baddeley, 1986; Ericsson and Chase, 1992; Masson and Miller, 1983). This correlation suggests that performance in the reading span test involves the use of LTM. In that case, it may be that the test measures the ability to store and retrieve information from LTM rapidly and accurately:

> We suggest that Daneman and Carpenter's (1980) reading span test measures this ability to store and later retrieve sentences about preceding sentences from LTM. Thus, what we are dealing with in the studies we have reviewed is not maintenance of temporary information in working memory, but skilled readers' ability to access LTM from retrieval cues held in the active portions of working memory (Ericsson and Kintsch, 1995: 228–229).

Similar considerations apply to Caplan and Waters' model of memory. These considerations are discussed below.

6.4.5 Implications for Caplan and Waters' model

Caplan and Waters' approach to human memory is based on Fodor's (1983) modularity hypothesis. According to this hypothesis, cognitive tasks are carried out by innate and special-purpose modules. Caplan and Waters propose that there are separate working memories for each module. This idea is somewhat implausible, because, taken at face value, it suggests that there are different modules for playing chess, diagnosing illness, playing basketball and so on. It seems more plausible to suggest that learned skills lead to the creation of domain-specific retrieval structures. The increased access to LTM made available by those retrieval structures can give the false appearance of there being separate working-memory capacities for different cognitive domains. As noted earlier, Fodor defines his modules in terms that are also true of learned skills. There is a strong possibility that the modularity hypothesis mistakes, to some degree, learned skills for innate modules.

6.4.6 Implications for MacDonald and Christiansen's model

Connectionist models are based on associative learning and they essentially implement the law of association by contiguity and the law of frequency. There is also a key connectionist principle which states that complex behaviour emerges spontaneously from the co-operative behaviour of simple processing elements. These principles do not appear adequate for the task of explaining how, for instance, memory experts can create hierarchical systems of cues through successive recoding of the material to be remembered. In addition, these principles do not shed light on how subjects can follow instructions to recall memorised digits in arbitrary orders. Associative learning is slow and incremental and this is difficult to reconcile with the flexible manner in which humans can follow arbitrary rules.

It has been suggested in previous chapters that the inability of connectionist models to support control processes is related to their inability to implement rule-governed processes. Ericsson and Kintsch's theory differs from connectionist modelling in that it accords a central role to conscious control processes. In theory, Ericsson and Kintsch's account could explain individual differences in the ability to use recursive rules. Individuals need to learn not only the structure of self-embedded sentences, but also how to encode such sentences in LTM in a manner which avoids interference between similar noun concepts. It has, in fact, been suggested by Hudson (1996) that the difficulty in comprehending self-embedded sentences is due to interference between noun concepts.

6.5 Conclusion

It needs to be noted, however, that Ericsson and Kintsch have little to say about syntactic structure. In order to apply their theory to syntactic processing, it is necessary to find a linguistic theory which is compatible with their general view of the memory system. In addition, Ericsson and Kintsch's theory gives a central role to task-specific systems of representation, but they do not provide a generalised description of such systems. The next chapter argues that Saussure's theory complements Ericsson and Kintsch's theory in these two respects. It will be shown that the theory integrates important aspects of rule-based and experience-based approaches to knowledge of language. In fact, that theory can be said to be the source of rule-based and experience-based approaches, which are simplifications of it. The combination of Ericsson and Kintsch's theory and Saussure's theory therefore provides a linguistically and psychologically motivated framework for describing individual differences.

7
Saussure's Theory of Language

7.1 Introduction

This chapter provides a description of Saussure's theory of language. According to this theory, the linguistic system in each individual's brain is constructed from experience. The process of construction depends on the associative principles of contrast, similarity, contiguity and frequency. The principle of contrast prevents confusion or interference between linguistic units by making them distinct from each other. The principle of similarity captures aspects of rule-governed behaviour by defining classes of intersubstitutable units which constrain combinatory processes. The principle of contiguity allows more complex units to be formed from the combination of simpler units. This principle also allows the products of the linguistic system to be integrated into the system in much the same way that mathematical formulae, which are the products of mathematical systems, become integral components of mathematical knowledge. The principle of frequency enables frequently occurring units to be recognised as stable units. It also enables abstract categories to be induced when a sufficient number of units turn out to have similar properties. It will be shown how the interaction of these principles enables the theory to incorporate rule-based and experience-based perspectives on language as well as giving the theory a certain psychological plausibility.

It should be noted that the theory is very general. It is not concerned with specific languages, such as English or French. In fact, it is not concerned specifically with language as such, but with systems of mental representation in general. Language is a case in point. In view of its level of generality, the theory does not make specific predictions about individual differences. It does, however, permit individual differences to arise in various ways.

It is necessary to bear in mind that the theory was put together from the lecture notes of Saussure's students. This, together with the fact that it has been translated from French, may be responsible for the fact that certain key points are not always treated with a desirable level of systematicity. For this reason, this chapter will combine description with interpretation. For the most part, the interpretation involves relating Saussure's ideas to some basic mathematical notions which are more clearly defined and easier to work with. This chapter begins by drawing some parallels between Saussure's theory and current ideas concerning the application of fractals to the biological modelling of plant growth.

7.2 A biological metaphor for the language system

Aspects of Saussure's theory are surprisingly contemporary; notably his commitment to the idea that complex structures emerge from the repeated application of simple rules. Currently, there is considerable interest in fractals, highly intricate mathematical objects that are generated by the repeated application of simple rules. Fractals display a property called self-similarity, whereby the parts display the same structure as the whole (see Mandelbrot, 1982). It turns out that formal linguistics has contributed indirectly to the study of fractals via Aristid Lindenmayer, who employed Chomsky's (1957) ideas on rewrite rules to model plant growth. To suit his purpose, Lindenmayer (1968) modified the manner in which rewrite rules are applied. His modifications turn out to be similar in some respects to the manner in which Saussure applied associative principles to describe the growth of language in the human mind. Lindenmayer's work is described very briefly below in order to provide a convenient metaphor of Saussure's theory.

Prusinkiewicz and Lindenmayer (1996) describe rewriting as 'a technique for defining complex objects by successively replacing parts of a simple initial object using a set of rewriting rules or productions'. They demonstrate the use of rewriting in the modelling of plant growth. According to Prusinkiewicz, plant forms are not really complex, but intricate:

> A plant is doing the same thing over and over again. Since it is doing it in many places, the plant ends up with a structure that looks complex to us. But it's not really complex; it's just intricate (cited in Devlin, 2000: 90).

When Prusinkiewicz says that a plant is doing the same thing repeatedly in many different places, he is alluding to the fact that plant growth can be modelled by applying rewrite rules in parallel. Chomsky (1957) applied rewrite rules serially to derive sentences from the initial symbol S. In order to model plant growth, however, Lindenmayer found it necessary for the same rule to be applied simultaneously to several parts of the plant. For instance, a rule which creates a branch has to be applied simultaneously to all parts of the plant where new branches can form. This mode of application is so powerful that it allows Prusinkiewicz and Lindenmayer (1996) to model the development of not only individual trees but gardens and even entire forests.

Prusinkiewicz and Lindenmayer's biological approach illustrates the shift in perspective needed to appreciate the status of rules in Saussure's theory. In Chomsky's (1957) theory, knowledge of language takes the form of a set of rewrite rules which are applied in order to generate individual sentences. In Saussure's theory, the rules do not comprise knowledge of language use; rather, they are principles for *constructing* language. Language itself is more like a growing plant and linguistic expressions are like branches. Just as new branches sprout from existing branches in accordance with basic patterns of plant growth, so, in Saussure's theory, new expressions sprout from existing expressions in accordance with basic principles of language growth. The linguistic system becomes progressively more intricate by incorporating its products into itself, just as new branches, fruits and leaves form integral parts of the tree.

7.3 Delimitation of the object of study

It is interesting to note that Saussure sets out his theory in a manner that is guided by the very principles which he holds responsible for the growth of language. This recursive manoeuvre will become apparent once the basic outline of the theory has been sketched. In hindsight, this recursiveness is only to be expected. Given that a theory is a representation, the mark of a truly general theory of representation is that it should be an instance of itself.

Saussure begins by identifying a number of linguistic dimensions: social (the communicative function), physical (the physical transmission of sound waves), physiological (the articulatory and auditory processes) and psychological (the association between sound representations and concepts). These dimensions comprise what he calls the speech circuit:

> Suppose, then, we have two people, A and B, talking to each other. The starting point of the circuit is in the brain of one individual, for

> instance, A, where facts of consciousness which we shall call concepts are associated with representations of linguistic signs or sound patterns by means of which they are expressed. Let us suppose that a given concept triggers in the brain a corresponding sound pattern. This is an entirely psychological phenomenon, followed in turn by a physiological process: the brain transmits to the organs of phonation an impulse corresponding to the pattern. The sound waves are sent from A's mouth to B's ear: a purely physical process. Next the circuit continues in B in the opposite order: from ear to brain, the physiological transmission of the sound pattern; in the brain, the psychological association of this pattern with the corresponding concept (1916: 12).

Saussure proposes that the primary object of linguistic study is the psychological dimension, which he refers to as linguistic structure: 'The linguist must take the study of linguistic structure as his primary concern, and relate all other manifestations of language to it.' (9). Having thus selected the relevant dimension for study, Saussure narrows the object of study further by defining it in contrast to alternative conceptions of language.

Firstly, he says that language is not innate: '. . . one may say that it is not spoken language which is natural to man, but the faculty of constructing a language . . . ' (10). This chapter will show that this faculty of constructing a language consists of the four associative principles of contrast, similarity, contiguity and frequency, though Saussure does not use these terms explicitly. Secondly, Saussure states that language is not 'a nomenclature: a list of terms corresponding to a list of things' (65). He objects to this idea because 'It assumes that ideas already exist independently of words'. This cannot be correct because 'If words had the job of representing concepts fixed in advance, one would be able to find exact equivalents for them as between one language and another.' (115). Thus the lack of exact equivalents between languages shows that linguistic representation is not veridical. Instead, it arbitrarily imposes order on otherwise amorphous sense data.

> Psychologically, setting aside its expression in words, our thought is simply a vague, shapeless mass. Philosophers and linguists have always agreed that were it not for signs, we should be incapable of differentiating any two ideas in a clear and consistent way. In itself, thought is like a swirling cloud, where no shape is intrinsically

> determinate. No ideas are established in advance, and nothing is distinct, before the introduction of linguistic structure (110).

7.4 Systems

The introduction of linguistic structure involves first of all setting up a system of contrasts which defines the basic units of language:

> Each language constructs its words out of some fixed number of phonetic units, each one clearly distinct from the others. What characterises those units is not, as might be thought, the specific positive properties of each; but simply the fact that they cannot be mistaken for one another. Speech sounds are first and foremost entities which are contrastive, relative and negative (117).

This point is easier to understand if it is cast in simple mathematical terms. The phonetic system of a language can be described as a set of dimensions and each phonetic unit takes a value along each dimension. This description does not capture the fact that units are defined in relative terms but it captures some important aspects of Saussure's view. For instance, languages select certain dimensions of the acoustic energy produced by the human vocal organs. These dimensions are defined in articulatory terms as voicing, manner of articulation, place of articulation and so on. Other acoustic dimensions, such as amplitude for instance, are ignored for the purpose of defining phonetic units. Each dimension has two or more values. For instance, the dimension of voicing has the values voiced and voiceless; the dimension of place of articulation has the values labial, dental, alveolar, velar and so on; the dimension of manner of articulation has the values plosive, fricative and so on. Languages can differ in the acoustic dimensions which they select and, where they select the same dimension, they can also differ in the number of values which they specify along that dimension.

Phonetic units are created by selecting one value from each dimension and combining the resulting set of values. For instance, /p/ is a voiceless labial plosive, while /b/ is a voiced labial plosive. Because each unit of speech sound has a value along each dimension, it is related in some way to every other speech sound: it is either similar to or different from every other speech sound along one dimension or another. The important point to note here is that phonetic units are the products of the overall system of dimensions and values. A small change to the system would result in changes to all the units of that system. For instance, if a system does not make a distinction between voiced and voiceless sounds, then there would be fewer speech sounds, because there would be no functional distinction between sounds such as /p/ and /b/ or /k/ and /g/ and so on.

The overall system of contrasts therefore simultaneously differentiates and groups phonemes: 'In linguistic structure everything in the end comes down to differences, and also to groups' (127). Differences between phonemes come about because each phoneme is a unique combination of values. Groups of phonemes come about in terms of shared values between phonemes. Thus, there are voiced versus voiceless phonemes, labials versus dentals versus alveolars and so on. Each phoneme will belong simultaneously to several groups, as many as there are dimensions.

More complex phonological forms are created by selecting phonemes from the phonetic system and combining them into larger units such as syllables. The selection and combination of phonemes to create syllables is sensitive to phoneme groupings. For instance, a syllable might consist of three phonemes, the onset, the nucleus and the coda. In that case, the onset and coda are consonants and the nucleus is a vowel. The combination of phonemes into syllables is therefore sensitive to phoneme groupings in that only phonemes possessing certain features may occur in certain positions in a sequence. This gives rise to rule-like behaviour: combinatory patterns are sensitive to classes of phonetic units rather than to individual phonemes.

It is important, at this point, to note that phonetic units are not defined in purely acoustic terms. Saussure states that, 'Linguistic signals are not in essence phonetic. They are not physical in any way. They are constituted solely by the differences which distinguish one sound pattern from another.' (117). To illustrate this concept, he introduces the concept of value and notes that 'it is not the metal in a coin which determines its value'. Thus, for instance, the same monetary value can be denoted by a coin, a note or a cheque. There are two interpretations of what Saussure means by this. On the one hand, he seems to be emphasising the fact that a unit of speech is not defined in terms of specific phonetic values, but rather in terms of a range of values defined by the overall phonetic system. On the other hand, Saussure also suggests that the acoustic correlates of a phoneme may be completely irrelevant and that what really matters is whether one phoneme can be substituted by another in a stretch of speech: 'In its place in a syntagma [sequence], any unit acquires its value simply in opposition to what precedes, or to what follows, or both.' (121). Saussure probably intended both interpretations, that is to say, phonemes are defined both in terms of phonetic contrasts and in terms of the contexts in which they occur.

To say that complex linguistic units are formed by combining simpler units is therefore slightly misleading. In actual fact, according to Saussure, the complex structures, to some degree, define their constituent units. This interdependence of part to the whole and the whole to the part is reminiscent of the modern notion of non-linearity. This notion captures the fact that there are certain phenomena where there is no clear distinction between cause and effect because effect contributes to cause and cause contributes to effect. In Saussure's theory, the parts create the whole and the whole creates the parts. This idea will be illustrated later in the chapter. For the purpose of simplifying the description, this chapter will maintain the convenient assumption that simple units are selected and combined to form progressively more complex units. It should be borne in mind, however, that things are not so straightforward for the language learner. It will be suggested that the attempt to sort out the wholes from the parts and vice versa is a major source of individual differences.

Continuing with the fiction of linearity, we can say that units of language that are more complex than syllables are created by selecting and combining syllables into morphemes (bearing in mind, however, that some morphemes, like plural –s, are shorter than syllables). Morphemes, in turn, are selected and combined to form composite forms, such as words. Words are selected and combined to form phrases, phrases into sentences, and so on. In each case, the same principles are at work: selection is based on the principle of similarity and combination is based on the principle of contiguity. Further, the outputs of each cycle of selection and combination become inputs to the next cycle.

According to Saussure, the same principles which give rise to progressively complex phonological forms are also responsible for the creation of progressively more complex concepts. Concepts can be thought of as combinations of selections from a set of more basic semantic features, just as words are combinations of selections from the system of syllables. Saussure illustrates the structure of concepts in terms of the following example:

> We assign identity, for instance, to two trains ('the 8.45 from Geneva to Paris'), one of which leaves twenty-four hours after the other. We treat it as the 'same' train, even though probably the locomotive, the carriages, the staff etc are not the same. [. . .] the train is identified by its departure time, its route, and any other features which distinguish it from other trains (107).

Concepts, like phonemes, are therefore defined in terms of dimensions and values along those dimensions. Each concept is a combination of values selected from the overall set of conceptual dimensions. Each concept is therefore related to every other concept in an intricate pattern of similarities and differences. As with phonemes, changes to the overall system will also result in changes in the number and nature of concepts. Systems of progressively more complex concepts are generated from systems of simpler concepts by repeated application of the principles of selection and combination. In addition, as indicated in the above quotation, different tokens can count as the same concept as long as they are interchangeable. Thus, even if two trains are completely different in physical terms, they count as the 'same train' if they occupy the same slot in the timetable. (Because of problems with the underground rail network, tube trains in London sometimes take the form of a bus!)

It is a measure of the generality of Saussure's view of language that the same principles of organisation govern the formation of both concepts and phonological forms. The generality goes further than this, because the same principles also govern the *relationship* between forms and concepts. This relationship is mediated by the linguistic *sign*, which is described in the next section.

7.5 The sign

A linguistic sign is formed when a form (signifier) and a concept (signified) are selected from their respective domains and combined. In a sense, the combination of form and concept is an instance of the more general operation of combining units from different groups or paradigms (in this case the combination is between units from phonological and conceptual domains). Once combined, a form and a concept form a single unit whose sub-units help stabilise each other: the form binds the bundle of conceptual values which comprise a concept and the concept binds the bundle of phonetic values which comprise the form:

> The characteristic role of a language in relation to thought is not to supply the material phonetic means by which ideas may be expressed. It is to act as an intermediary between thought and sound, in such a way that the combination of both necessarily produces a mutually complementary delimitation of units (110).

The effect of this interdependence between form and concept is to make signifier and signified inseparable:

> A language might also be compared to a sheet of paper. Thought is one side of the sheet and sound the reverse side. Just as it is impossible to take a pair of scissors and cut one side of paper without at the same time cutting the other, so it is impossible in a language to isolate sound from thought and thought from sound (111).

This is yet another example of the 'non-linearity' described earlier. On the one hand, the signifier and the signified create the sign. On the other hand, the sign creates the signifier and the signified in the sense that it is by virtue of the sign that the bundle of phonetic features comprising the signifier and the bundle of conceptual features comprising the signified are bound together and stabilised as unitary constructs. This is not to say that there can be no phonological forms of concepts independently of the sign, simply that the stability of the identities of such forms or concepts depends on their signifying relationship.

Having created unit signs, the operations of selection and combination create more complex signs from simpler ones: '...we do not express ourselves by using single linguistic signs, but groups of signs, organised into complexes which themselves are signs.' (127). This quotation indicates that Saussure considers the concept of sign to be a recursive one: a combination of signs is itself a sign. The next section shows that 'sign complexes' are organised in ways consistent with the principles of similarity and contiguity.

7.6 Complex signs

In describing the relationships between signs, Saussure states that:

> The relations and differences between linguistic items fall into two quite distinct kinds, each giving rise to a separate order of values. The opposition between these two orders brings about the specific character of each. They correspond to two different forms of mental activity, both indispensable to the workings of language (121).

The 'two forms of mental activity' are, of course, the operations of selection and combination which are based respectively on the principles of similarity and contiguity. The process of selection operates on items which are similar along some dimension and the process of combination combines signs into more complex signs. Saussure expands on the nature of the two types of relationships as follows:

> Considered from these two points of view, a linguistic unit may be compared to a single part of a building, e.g. a column. A column is related in a certain way to the architrave it supports. This disposition, involving two units co-present in space, is comparable to a syntagmatic relation. On the other hand, if the column is Doric, it will evoke mental comparison with the other architectural orders (Iconic, Corinthian, etc.), which are not in this instance spatially co-present. This relation is associative (122).

The syntagmatic relation, which involves 'two units co-present in space', is created by the combinatory process. The selective process, on the other hand, operates on the associative relation, which involves similar items that are not 'spatially co-present'. The term 'associative' is rather confusing, given that syntagmatic relations also involve (sequential) association. To avoid possible confusion, the term 'paradigmatic', favoured by Roman Jakobson, will be used instead. Each sign in the language system is therefore linked to other signs paradigmatically and syntagmatically. In mathematical terms, this idea translates into that of a graph – a collection of nodes linked by paradigmatic and syntagmatic connections. The following sections describe the nature of paradigmatic and syntagmatic connections respectively.

7.7 Paradigmatic connections

The discussion of phonemes showed how the overall system of dimensions and values differentiates and groups phonemes with respect to the combination of values which defines each phoneme. Phonemes that share a value in common belong to a group identified in terms of that value (for example, labials, fricatives, plosives and so on). Furthermore, because each phoneme is a combination of values selected from several dimensions, each phoneme will belong to several groups. The same grouping principle applies to signs, except that here the grouping criteria are less constrained by the internal structure of each unit:

> Groups formed by mental association do not include only items sharing something in common. For the mind also grasps the nature

of the relations involved in each case, and thus creates as many associative series as they are relations (124).

Signs can be connected paradigmatically on the grounds of any possible association between them: 'Any word can evoke in the mind whatever is capable of being associated with it in some way or other.' (124). Saussure illustrates this point with a graph (Figure 7.1).

The first series of associations is related by the concept of 'teaching'; the second by the concept of 'learning'; the third by the suffix '-ment' and the fourth by the purely phonological similarity in the final syllable '-ent'. The third series is actually related in two ways, grammatically by the suffix '-ment' and phonologically by the similarity in the final syllable, which is also '-ment'. Saussure notes: '... sometimes there is a double associative link based on form and meaning, but in other cases just one associative link based on form or meaning alone' (124). Paradigmatic connections have 'indeterminate order and indefinite number' (124) and each item 'acts as the centre of a constellation, from which connected terms radiate ad infinatum.' (124).

It is not clear whether Saussure thought that paradigmatic connections are formed voluntarily or involuntarily. On the one hand, he says that 'Any word can evoke in the mind whatever is capable of being associated with it in some way or other.' (124) and 'it is impossible to say in advance how many words the memory will suggest, or in what order.' (124). These two quotations make it sound as though paradigmatic

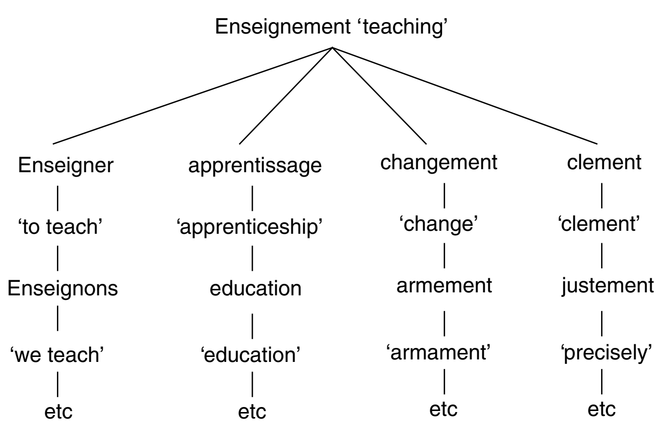

Figure 7.1 Paradigmatic connections

connections are formed automatically and without conscious volition. On the other hand, in discussing the formation of grammatical categories, he states:

> The three Latin genitives domini (of a master), regis (of a king) and rosarum (of a rose) have three different endings, -i, -is and –arum, which afford no phonetic basis of association. Nonetheless, these endings are linked by one's awareness of their common value, which prescribes identical uses for them. That is sufficient to set up an association, in the absence of any material support. And that is how the notion of 'genitive', as such, takes its place in the language (136).

This quotation suggests that the language user consciously induces abstract categories by noticing similarities between signs. Induction depends on there being a sufficient number of elements with which to form a category and therefore depends on the law of frequency.

In talking about complex signs, Saussure appears to be talking about the combination of signs into more complex structures. However, there is another sense of 'complex sign' and it is hard to tell whether or not Saussure intended it. This is the sense whereby the signified of one sign is not a simple concept or a combination of concepts, but a *set* of other signs. For instance, the 'notion of genitive' can be thought of as a complex sign in the sense that it signifies the set of signs which share the feature of being genitives. Saussure's example of columns suggests this interpretation, as argued below.

The signified of 'column' is presumably the set of signs each of which signifies an individual type of column, such as Doric, Iconic and so on. In other words, the sign 'column' captures the general notion of 'columness' and leaves out the specific details which identify a particular column as Doric, Iconic and so on. In other words, the sign 'column' provides an underspecification of the concept of column, which is 'inherited' by specific types of column. This observation applies to any general term, for instance, animal, car, country and so on which can be broken down into subtypes. If Saussure intended this interpretation of 'complex sign', then it would suggest that the linguistic system can develop taxonomic hierarchies in terms of signs of signs in a potentially unlimited chain of abstraction. If so, then the potential for rule-governed behaviour is limitless because there would be no limit to the abstractness of mental categories.

It is interesting to note that the notion of rule can be reduced to a paradigmatic phenomenon. Saussure himself does not suggest that language users employ abstract grammatical rules. However, in principle, his theory permits such rules to emerge. A rewrite rule such as S → NP VP is an instruction to rewrite the symbol S as the symbol string NP VP. The relationship between S and NP VP is a paradigmatic one in the sense that S can be seen as a paradigm, and NP VP as an element of that paradigm, together with other ways of rewriting S, such as AdvP NP VP. In mathematical terms, one could speak of S as a set and NP VP as one member of that set. By the same token, NP can be regarded as a paradigm, among whose elements are all the possible ways of rewriting NP, such as Det N, Det AdjP N, N S and so on. The same goes for VP.

In that case, a rewrite rule is an instruction to select one member of a given paradigm. Thus S → NP VP can be seen as an instruction to select the string NP VP from the paradigm S. The rule NP → Det N can also be seen as an instruction to select the string Det N from the paradigm NP, and so on. In that case, a rewrite grammar can be seen as a set of instructions to make certain selections from certain paradigms. The process of selection would ultimately produce a string of words, or sentence. Alternatively, a rewrite grammar need not be seen as a set of rules at all, but as a description of the hierarchy of paradigms to which words in a sentence belong. In theory, therefore, a rewrite grammar can be reduced to the operation of the law of similarity. As noted earlier, Saussure does not discuss formal rules and the preceding discussion is an extrapolation of the theory.

7.8 Syntagmatic connections

The discussion on phonemes showed that phonemes are combinations of choices. Similarly, signs are combinations of signifiers and signifieds. Further, signs can also be combined to form more complex signs:

> Word as used in discourse, strung together one after another, enter into relations based on the linear character of words [...] Combinations based on sequentiality may be called *syntagmas* (170).

For instance, a sentence is a complex sign formed through syntagmatic connections between its constituent signs:

> ... the notion of a syntagma applies not only to words, but to groups of words, and to complex units of every size and kind (compound words, derivative forms, phrases, sentences) (122).

Saussure notes that syntagmas vary in the degree to which they are fixed or formed by the free combination of their constituents. On the one hand,

> There are [...] a large number of expressions belonging to the language: these are ready-made phrases, absolutely invariable in usage, in which it may even require reflection to distinguish the constituent parts...(123).

'These oddities' Saussure continues, 'are not improvised, but handed down by tradition ... ' (123). He gives examples such as, 'what's the use?'; 'come along'; 'to take offence'; 'to force someone's hand' and so on. The idea is that there are certain complex signs which are rote-learned.

Novel syntagmas, however, are formed by modifying existing forms.

> When a new word, such as indecorable ('undecoratable') crops up in speech, it presupposes a pre-existing type, and the type in question would not exist were it not for our recollection of a sufficient number of similar words already in the language (123).

Thus new words are created by analogy to existing words. It may be noted here that the law of frequency is operative here: a sufficient number of precedents is required for a new combination to be made. The same principle applies to the formation of novel sentences:

> Exactly the same holds for sentences and groups of words based upon regular models. Combinations like...'the earth rotates'... 'what did he say to you?' etc correspond to general combinatory types, which in turn are based in the language on specific examples heard and remembered (123).

The quotations are the basis for the biological metaphor of the linguistic system which was suggested at the beginning of this chapter. New forms are derived from existing forms just as branches sprout from existing branches. An individual who has heard the sentence 'the earth rotates' can, in this view, create new sentences by substituting the words in the sentence for other words, for example, 'the earth revolves'; 'the wheel rotates' and so on.

'General combinatory types' are complex signs in two meanings of the expression. Firstly, they are complex signs in the sense of being syntagmas, that is, combinations of signs. Secondly, they are complex signs in the sense that they group together a set of sentences (themselves complex signs) which share certain features in common. Earlier, the sign 'column' was described as signifying an underspecified concept of 'column-ness'. By the same token, a sentence type is a complex sign which underspecifies a set of sentences. For instance, the general combinatory type to which the sentence 'The earth rotates' belongs to might specify a sequence of parts of speech such as Determiner Noun Verb. Parts of speech are themselves abstract types, being signs of signs. In that case, a sentence type is a syntagmatic connection between signs of signs.

There is a mathematical parallel to this form of creativity. An algebraic expression, such as the one which says that the area of a circle is equal to pi multiplied by the square of the radius, specifies certain constant and variable terms in a fixed relationship. One can use this expression to calculate the areas of specific circles by replacing the place-holder (radius) with a specific numerical value for the radius. General combinatory types can therefore be described as formulae in a sense that is akin, but not identical to, mathematical usage. In addition to supporting formulaic uses of language, Saussure's theory also allows for formal derivational processes, as suggested earlier. Saussure did not develop the theory in that direction. The point to note here is simply that his theory is capable, at least embryonically, of integrating both associative and formal aspects of language. The issues raised in the last few paragraphs are central to understanding differences between rule-based and experience-based approaches to language and to understanding how Saussure's theory integrates the insights of each approach while avoiding the problems associated with each. It is therefore worthwhile to take a brief detour at this point in order to explore these issues in more detail.

7.9 A brief detour

Chomsky (1967) is highly critical of the idea of sentence patterns based on sequences of parts of speech. This criticism is important because it represents a key objection to the idea that syntactic structure can be induced from experience. It is therefore important to deal with it in some detail in order to show that the objection is not compelling. Chomsky (1967: 400) states:

> The idea that a person has a "verbal repertoire" – a stock of utterances that he produces by 'habit' on an appropriate occasion – is a myth, totally at variance with the observed use of language. Nor is it possible to attach any substance to the view that the speaker has

> a stock of "patterns" in which he inserts words or morphemes. Such conceptions may apply to greetings, a few clichés, and so on, but they completely misrepresent the normal use of language.

Although Chomsky is dismissive of the idea of sentence patterns, it is worthwhile to consider the co-existence of derivational rules and formulae in mathematics, as suggested above. For instance, Pythagoras' theorem is the product of a deductive process. However, once that expression was obtained, it has been used subsequently by simply inserting values for the known variables in the formula in order to obtain a value for the unknown variable. The point of this example is that there is no necessary contradiction between the existence of rules and the existence of formulaic expressions. It actually makes sense to store the products of computation in order to facilitate future computations. Another important observation follows from this.

Schoolchildren can learn and use mathematical formulae successfully without any knowledge of how those formulae were derived. There seems to be no good reason to exclude the same possibility in the use of language. It may be that there are some individuals, who, like professional mathematicians, can derive complex expressions in their native language through verbal reasoning. There may also be other individuals, who, like the schoolchildren in the example, simply treat complex expressions in the language as formulae, into which they insert lexical variables.

It is interesting that Saussure treats expressions containing subordinate clauses as fixed types. He considers, for instance, that the expression 'the man I have seen' is based on a general combinatory type. For Chomsky (1957), subordination is effected by recursive rules. However, as shown in Chapter 3, the experimental evidence shows that, while subjects can process sentences with one degree of self-embedding, a significant number have difficulty with two or more levels of self-embedding. It may be that such subjects represent subordinate clause expressions such as 'the man I have seen' as combinatory types. This form of representation would explain the failure to generalise to two levels of self-embedding.

Chomsky (1965) suggested that language users are not necessarily aware of certain patterns in their native language. He illustrated this idea with reference to awareness of ambiguity by saying that language users are not necessarily aware of the triple ambiguity in the sentence, 'I had a book stolen'. He proposed that this ambiguity can be brought to a language user's awareness by paraphrasing the sentences in three different ways: (i) 'someone stole a book from my car'; (ii) 'I had someone steal a book' and (iii) 'I had almost succeeded in stealing a book'. Chomsky (1965: 21–22) proposed that:

> 'In bringing the triple ambiguity to consciousness . . . we present no new information to the hearer and teach him nothing new about his language but simply arrange matters in such a way that his linguistic intuition, previously obscured, becomes evident to him'.

Freedle and Craun (1970) suggested that Chomsky's argument above can be applied to the comprehension of self-embedded sentences. That is to say, language users possess recursive rules, but for some reason, this knowledge is obscured. Freedle and Craun therefore wanted to see if providing subjects with examples of doubly self-embedded sentences would enable them to recall their knowledge of recursive rules. It was found that examples indeed helped subjects understand the test items. However, some subjects did not benefit from the examples.

An interesting result from Freedle and Craun's experiment is that subjects shown examples of sentences containing two self-embeddings could generalise to test sentences with two levels of self-embedding but not to sentences containing three levels of embedding. If the examples used by Freedle and Craun did not tell their subjects anything new about English, but simply clarified their previously obscured intuitions, as Chomsky suggests, then it would have to be concluded that these subjects possessed grammars which specified that self-embedding does not exceed two levels. This suggestion is rather implausible and it seems much simpler to suggest that subjects *induced* a rule from the examples given. It could be that subjects differed in inductive ability, such that some subjects induced the recursive rule correctly; others did not induce the rule to a sufficient level of generality and others did not induce the rule at all.

The preceding discussion brings to light a certain problem in language learning. Each learner is confronted with a variety of expressions in the language. Some of these expressions are regular, in the sense that they can be derived from more basic principles (the core grammar in current linguistic theory). Other expressions, however, are idiomatic, being either completely frozen or exhibiting varying degrees of productivity. The learner has no way of knowing in advance which expressions are regular and which are peripheral. In Saussure's theory, language is learned by induction from experience. Given that individuals appear to differ in inductive ability, it is possible that individuals will differ in their ability to define appropriate units and categories of units from their linguistic experience. The next chapter will present some evidence which supports this possibility.

Saussure is also able to explain a class of syntactic phenomena which cause serious problems for rule-based approaches. This class of phenomena relates to context effects. In making the argument for the rule-based approach, Fodor and Pylyshyn (1988) state that syntactic expressions are compositional. That is to say, each lexical unit makes the same semantic contribution in different contexts. They give the example 'x is a brown cow' and state that the expression can be broken down to the conjunction 'x is brown' and 'x is a cow'. However, there are many expressions in English which cannot be broken down in this manner, indicating that their structure is not compositional. For instance, 'x is a great loser' does not break down to 'x is great' and 'x is a loser', nor does 'x is a giant ant' break down to 'x is a giant' and 'x is an ant'. These expressions are interesting examples of non-linearity in language: the adjective modifies the noun and the noun modifies the adjective. The whole is not simply the sum of its parts: the whole also defines the parts.

Saussure's notion of combinatory types is consistent with Construction Grammar, an approach to grammatical description which deals with the context-sensitive aspects of language. As will be shown below, this aspect of language can only be handled awkwardly using rules, by postulating a multiplicity of lexical items which correspond to each structural context. In Construction Grammar, constructions form contexts which modify the meanings of words. Goldberg (1996) illustrates this point with the aid of the *way*-construction (for instance, 'she sewed her way to success'). According to Goldberg, this construction is a productive schema with the general form: NP + V + Possessive + *way* + Oblique object. The choice of lexical filler in each slot is constrained to a greater or lesser extent. Some lexical elements in the schema are invariable, such as 'way'. Slightly less constrained is the possessive, which, however, must agree with the subject NP. The oblique object is constrained to adverbial directionals. The least constraint relates to the requirement that the first lexical item(s) comprise an NP followed by a Verb. Goldberg discusses some subtle constraints on the choice of the types of verb which can enter the construction.

Goldberg shows how the schema influences the meaning of a verb which is inserted into the verb slot. Examples of the way-construction are:

1. Frank *found* his way to New York.
2. Nitya *sewed* her way to fame and fortune.
3. He *belched* his way out of the restaurant.

The feature of interest with this construction is that the verbs 'found', 'sewed' and 'belched' do not normally convey a sense of motion, yet they are construed in this manner in the examples above. There are two main accounts of this effect. Levin and Rapoport (1988, cited in Goldberg, 1999) suggest that verbs which appear in the way-construction have more than one meaning, and one of those meanings relates to motion. Effectively, what Levin and Rapoport propose is that there is special sense of 'found' which is 'to move by finding', a special sense of 'sewed' which is 'to move by sewing' and a special sense of 'belch' which is 'to move by belching'. The attempt to explain context-effects by multiplying the number of lexical items in this way is common in rule-based approaches to linguistic description. Its attraction is that linguistic structure can then be explained in terms of context-free rules.

However, Goldberg (1999) argues that such an account requires an implausibly large number of word-senses. She offers numerous examples from the OUP corpus which show that the way-construction is highly productive:

1. He'd *bludgeoned* his way through . . .
2. [The players will] *maul* their way through the middle of the field.
3. Their customers *snorted and injected* their way to oblivion and sometimes died on the stairs.
4. But he consummately *ad-libbed* his way through a largely secret press meeting.
5. I cannot inhabit his mind or even *imagine* my way through the dark labyrinth of its distortion.
6. Lord King craftily *joked and blustered* his way out of trouble at the meeting.

Given the enormous number of verbs which participate in this construction, Goldberg suggests that it is descriptively more economical to regard the sense of motion as a property of the way-construction, rather than as a property of each verb. Verbs only acquire the motion sense in the context of the construction. Accepting this view means that constructions need to be regarded as linguistic entities in their own right, as opposed to being mere epiphenomena.

The preceding discussion shows that Saussure's theory contains the best of both worlds. The unlimited potential for abstraction allowed for by the theory enables it to capture rule-governed aspects of linguistic behaviour. By permitting memory of previously experienced expressions to influence the production and comprehension of new expressions, the theory is able to capture context-effects. The next section considers how the combination of paradigmatic and syntagmatic connections come together to form the image of the language system as graph.

7.10 The language system as a graph

Saussure states that, 'In a linguistic state, then, everything depends on relations' (121). The previous sections showed how linguistic units are related to each other via syntagmatic and paradigmatic connections. This is true of linguistic units at every scale. For instance, phonemes are formed through combinatory links between phonetic features. Each phoneme is also connected paradigmatically to every other phoneme with which it shares a common feature. Syllables are formed from combinatory links between phonemes and each syllable is connected paradigmatically to every other syllable with which it shares a phoneme. Words are formed from combinatory links between syllables and are connected paradigmatically with every other word with the same syllables or with other words with the same features of meaning. The same is true for phrases, clauses and sentences. Saussure states:

> The whole set of phonetic and conceptual differences which constitute a language are thus the product of two kinds of comparison, [paradigmatic] and syntagmatic. Groups of both kinds are in large measure established by the language. This set of habitual relations is what constitutes linguistic structure and determines how the language functions (126).

It is also interesting to note that, when Saussure compares different languages, each language is characterised as a combination of choices. For instance, Saussure notes that: '...a distinction could be drawn between *lexico-logical* languages [...] and *grammatical* languages' (132). The former uses separate lexical items to make certain distinctions, while the latter uses grammatical devices for the same purpose. Languages can therefore be said to differ along the lexico-logical/grammatical dimension, with each language taking one value or the other along this dimension. Although Saussure does not dwell on this point, the idea that languages differ along several dimensions is an interesting one and it has been developed extensively by Chomsky in terms of a universal grammar. The putative universal grammar is said to provide, among other things, a set of choices along several dimensions or 'parameters'. Each language is defined by the combination of choices made from the universal grammar. In that case, each language is defined by combinatory links between values from the universal set of choices and paradigmatic links to other languages which make similar choices.

Thus there is a sense in which the structure of the entire language system is the same as the structure of its smallest unit, the phoneme: it is essentially a combination of choices. This structure can be represented by a graph. This graph consists of nodes (representing linguistic units) which are connected via syntagmatic and paradigmatic links to other nodes. Units at each level of linguistic structure can be represented using nodes with syntagmatic and paradigmatic connections. Even an entire language can itself be represented in this way. The recursion of the same structure at different levels of structural organisation is a linguistic instance of self-similarity.

There is an alternative form of representation which also exhibits a fractal structure. As stated above, the units of a language, including the language itself, can be described as combinations of values along several dimensions. We can think of these values as defining multidimensional space, or hyperspace (recall Elman's (1992) analysis of the representational space formed by hidden unit activation values). In that case, each linguistic unit is represented by a point in a hyperspace. This would be true of units at all levels of structural organisation, given that they have the same structure. An entire language could, in theory, be represented as a point in the hyperspace of languages.

The next section describes how paradigmatic and syntagmatic connections relate to Saussure's distinction between language and speech.

7.11 *Langue* and *parole*

The discussion so far has focused on the language system implemented in the brain. The discussion has shown how the system grows by recycling the output of selective and combinatory processes back into the same processes at higher levels of structural organisation. These processes are carried out during individual acts of speech. That is to say, the language user selects and combines units in order to produce or understand utterances. The output of these processes is transmitted into the social environment and simultaneously stored in memory. The transmitted output of all language users comprises a public corpus of expressions, of which the individual user possesses a sample in LTM. Saussure makes a distinction between the associative linguistic memory, which he calls *langue*, and the public corpus, which he calls *parole*:

> Speech is the sum total of what people say, and it comprises (a) individual combinations of words, depending on the will of the speakers, and (b) acts of phonation, which are also voluntary and are necessary for the execution of the speakers' combination of words (19).

There is a close relationship between syntagmatic connections in the brain and the expressions in speech. Syntagmatic connections represent the combination of signs selected in particular instances of language use. The collection of syntagmas in the brain therefore corresponds to the corpus of expressions in speech in such a way that it is difficult to distinguish clearly between them:

> Where syntagmas are concerned [...] one must recognise the fact that there is no clear boundary separating the language, as confirmed by communal usage, from speech, marked by freedom of the individual. In many cases it is difficult to assign a combination of units to one or the other. Many combinations are the product of both, in proportions which cannot be accurately measured (123).

The case is different, however, with paradigmatic connections. These exist only in the brain (that is, in the language system) and cannot be perceived directly in speech. However, paradigmatic connections are *latent* in speech and can be discovered by attempting to systematise speech. If paradigmatic connections were not latent in speech, then, according to Saussure, it would be impossible to learn language:

> These two objects [language and speech] are doubtless closely linked to and each presupposes the other. A language is necessary in order that speech should be intelligible and produce all its effects. But speech is also necessary in order that a language may be established. Historically, speech always takes precedence. How would we ever come to associate an idea with a verbal sound pattern, if we did not first of all grasp this association in an act of speech? Furthermore, it is by listening to others that we learn our native language. A language accumulates in the brain only as a result of countless experiences. Finally, it is speech which causes a language to evolve. The impressions received from listening to others modify our own linguistic habits. Thus there is an interdependence between the language itself and speech (19).

The language learner therefore has to reconstruct language in his or her own brain by subjecting speech data to the recursive processes of selection and combination. Saussure does not take a clear view on whether the resulting system is identical for all learners of a language. On one hand, he suggests that the grammatical systems of all users of a language are the same:

> All individuals linguistically linked in this manner (i.e. socially) will establish among themselves a kind of mean; all of them will reproduce – doubtless not exactly, but approximately, the same signs linked to the same concepts (13).

And,

> A language, as a collective phenomenon, takes the form of a totality of imprints in everyone's brain, rather like a dictionary of which each individual has an identical copy (19).

Elsewhere, however, he says that 'language is never complete in any single individual, but exists perfectly only in the collectivity.' (13). He suggests that the 'copy' in each brain differs in some ways from the 'copies' in other brains.

7.12 Conclusion

A reference was made at the beginning of the chapter to the fact that there is a recursive aspect to the way that Saussure sets out his theory: he employs the very same principles that he is about to describe. It may be recalled that Saussure begins by isolating various dimensions of language and *selecting* one of them for special attention. He then defines his concept of language by *contrasting* it with alternative theoretical conceptions of language: it is constructed rather than innate and it represents reality in a selective rather than veridical manner. In other words, Saussure constructs his theory of language by *combining* theoretical choices from a set of alternatives. A detailed description of his theory would show a complex pattern of similarities and differences with other theories. His theory could therefore be described, in relation to other theories of language, as a point in the hyperspace of linguistic theories.

The process of constructing a theory of language therefore mirrors the process by which language itself is constructed by each individual. As stated at the beginning of the chapter, this recursiveness is to be expected, if only in hindsight: a sufficiently general theory of representation, being itself a representation, has to be an instance of itself.

8
Patterns of Individual Differences

8.1 Introduction

This chapter argues that language users construct different linguistic systems and consequently, differ in the manner in which they define their linguistic units. These differences lead to individual variations in linguistic representation. This argument is based on an application of Saussure's theory to data on individual differences reported in the experimental literature. The theory describes the construction of the language system in terms of four associative laws: the laws of contrast, similarity, contiguity and frequency. The chapter applies the theory by treating each law as a dimension of individual differences. No attempt is made to delve into the ultimate reasons why individuals differ in their application of the four laws. The chapter begins with a caveat about the application of Saussure's theory to individual differences.

8.2 The four laws of association

It is important to bear in mind that Saussure's theory is a linguistic rather than a psychological theory. It is therefore not clear how best to apply it to experimental data. The chapter takes a conservative approach by using the theory in a purely descriptive manner. It considers the sorts of individual differences that might logically be expected to arise from variations in the application of the four associative laws. It then treats these laws as dimensions of individual variation and considers whether specific individual differences that have been reported in the literature can be categorised in terms of the application of one law or another. The laws interact, making it difficult to assign specific individual differences to particular laws. Assignments are therefore made on the basis of the law which seems particularly salient in each case. This discussion provides a useful background to the experiment described in the next chapter. The experiment was carried out to decide if native speaker variations in understanding complex sentences are due to variations in working-memory capacity (Just and Carpenter, 1992) or to variations in syntactic representation (Ericsson and Kintsch, 1995). While the experimental hypothesis was not derived from Saussure's theory, the theory fits the experimental data well.

According to the theory, linguistic units are defined contrastively. This is the law of contrast at work. For instance, phonemes are defined in contrast to other phonemes. However, the pattern of contrasts gives rise to similarities between phonemes which take the same value along some dimension. These similarities define paradigms or classes of phonetic units. This is the law of similarity in operation. There is therefore a close relationship between the law of contrast and the law of similarity. Phonemes are also defined in terms of their combinatorial possibilities, so that two phonemes which are phonetically distinct but have the same combinatorial behaviour can be treated as the same phoneme. Thus the law of contrast and the law of contiguity are also closely related. Furthermore, patterns of combination are defined in terms of paradigms or categories of units, indicating a close relationship between the laws of similarity and contiguity.

Finally, the law of frequency is related to the other three laws. It is related to the law of similarity because similarities can only be observed if they apply to a number of units. To put this differently, linguistic categories are induced only when a sufficient number of items are observed to share some linguistically relevant feature. It would also seem to be the case that contrasts become more salient in relation to the number of cases to which they apply. Thus the laws of frequency and contrast interact. The law of frequency interacts with the law of contiguity in at least two ways. Firstly, the law of frequency concerns the probability that one unit will occur after another in a sequence. Secondly, frequently co-occurring units can become unitised, thus forming a single perceptual or conceptual unit. Conversely, a unit could be analysed into sub-units if those sub-units occur across a wide range of other units.

The generalisations made above derive from a logical extrapolation of Saussure's theory to individual differences. The next paragraphs discuss experimental data to see how well these generalisations apply. The discussion covers individual differences in (a) phonology; (b) morphology; (c) lexicon and (d) syntax.

8.3 Phonology

Bates *et al.* (1988) summarise several reviews of the literature on individual differences in first language development. They note various dichotomies in the ways that children learn the phonological systems of their native language. According to Bates *et al.*, these dichotomies do not represent types of children, rather, they are extremes with continuous variation in between.

In one dichotomy, one group of children tends to be word-oriented and to emphasise segmental aspects of phonology. Another group, however, is oriented towards intonation. They emphasise suprasegmental aspects of phonology and their speech contains 'filler sounds' which make it sound adult-like. This pattern of differences can be described in terms of the law of contiguity in the sense that it relates to differences in the size of basic phonological units. Another dichotomy relates to the consistency of pronunciation across word tokens: one group of children displays consistency across word tokens while another displays variations across word tokens. The latter group appears to possess more context-dependent phonological representations. It could be that this group possesses phonological chunks which have not been sufficiently analysed into smaller, context-free units. This pattern of differences could also be assigned to the law of contiguity in as much as it relates to the size of phonological units. Children also vary in terms of intelligibility. Some children have highly intelligible speech while others have comparatively unintelligible speech. This pattern of differences can be described in terms of the law of contrast. The latter group of children appears not to make clear distinctions between different units of speech.

It appears that some of the differences described above persist into adulthood. Day (1969, 1979) found that adults differ in terms of bottom-up or top-down processing during speech perception. Using a dichotic listening task, she presented BANKET to one ear, 50–150 ms before or after presenting LANKET to the other ear. Subjects were asked to report whether they heard B or L first. One group of subjects reported hearing the correct phoneme sequence whichever word was presented first. Other subjects reported hearing B first, even if LANKET was presented before BANKET. The first group of subjects appeared to be performing the task in a bottom-up fashion, whereas the second group appeared to perform the task in a top-down fashion. This interpretation suggests that the subjects differed in perceptual bias. Some subjects seemed to pay greater attention to phoneme units in the stimuli, while other subjects seemed to pay greater attention to lexical units and this led them to make errors. The individuals therefore differ in the size of unit which they attend to. Paying attention to lexical units can result in loss of precision in the perception of the sub-units.

There is also evidence of an interaction between contiguity and frequency. Chapter 3 referred to research on reading, which shows that skilled readers are more sensitive to transitional probabilities in orthography than less skilled individuals (for example, Lefton *et al.*, 1973; Scheerer-Neumann *et al.*, 1978). In other words, child readers differ in their knowledge of the combinatorial probabilities of orthographic units. De Fazio (1973) found that adults also differ in the ability to predict verbal sequences. This general pattern of differences appears to be related to the frequency with which certain sequences occur. It could be that these differences arise from differences in linguistic experience.

8.4 Morphology

Berko (1958) used inflected pseudo-words to study the development of inflectional morphology in English speaking children. Given a word like 'wug', the children could, under appropriate circumstances, produce the plural – 'wugs'. This study was interpreted to show that children are aware of morphological rules of plural formation. Schnitzer (1993) used a similar technique to study knowledge of Spanish verbal inflectional morphology in Spanish children (7–11 years old) and adults (at least 18 years old). He asked his subjects to perform a conjugation task. The task involved filling in the gaps in sequences such as 'you dummy verb + inflection; we dummy verb + inflection; I ____.' Surprisingly, Schnitzer found that the children performed the task better than the adults.

Schnitzer suggests that children and adults differ in their categorisation of novel forms. He argues that, in Spanish, new coinages are assigned to the first conjugation. Adult speakers might therefore respond to a novel form by treating it like a new coinage. Children, however, appear not to categorise novel forms as coinages. It could be that the adults have a larger vocabulary of coinages which forms an analogical basis for treating novel forms as coinages. The pattern of differences between children and adults can be described in terms of differences in applying the law of similarity. The adults appear to treat pseudo-words as coinages and therefore inflect them like coinages. Children, on the other hand, appear to inflect the pseudo-words in terms of regular Spanish morphosyntax. Thus differences in categorisation affect combinatorial patterns.

8.5 Lexis

There is evidence that both children and adults differ in lexical distinctions. That is to say, some individuals appear not to distinguish between contexts in which certain lexical items should be used. Maratsos (1976) tested English children and adults on the use of noun definitisation in noun phrases referring to a specific but non-unique referent. For instance, if the hearer does not know that the speaker has a hammer in her bag, then it would be appropriate for the speaker to use the indefinite article in referring to the hammer, for example, 'I have a hammer in my bag'. On the other hand, if the hearer knows about the hammer, then the definite article would be appropriate, for example, 'I have the hammer in my bag'. The distinction between the definite and indefinite article in such contexts depends on the speaker's awareness of what the hearer knows. Maratsos found that about half of the children and half of the adults in his study did not make this distinction correctly. He concluded that 'perhaps competence, at least as measured by actual performance of the above kind, never reaches a perfect state' (1976: 105, cited in Biber, 1983). In this case, the pattern of differences relates to the law of contrast.

Biber (1983) also found individual differences in knowledge of a noun definitisation rule among speakers of a Somali dialect (different from the English rule studied by Maratsos). Biber suggests that knowledge of the rule which he studied depended on whether or not an individual had heard oral narratives in childhood. Biber (1983: 293) suggests that '...it is possible for mature adults to lack "competence" in certain portions of the grammar of their own native tongue, simply due to lack of adequate exposure to these portions of the grammar.' This pattern of difference appears to be related to frequency, in the sense that individuals appear unable to understand rules which do not occur in their experience.

Cupples and Holmes (1992) found that native English speakers can differ in knowledge of lexical classes. They asked university students to make yes-no judgements on whether pairs of words could perform the same function in a sentence. For instance, *sailor-theft* can perform the same syntactic function but not *illness-diminish*. Subjects were also asked to perform a semantic relatedness judgement. For instance, *steal-theft* are semantically related but not *wisdom-obsolete*. They found that skilled readers performed better than less skilled readers in the syntactic judgement task but there were no differences between the two groups on the semantic judgement task. Differences in lexical categorisation are related to the law of similarity. As suggested earlier, such differences can be expected to affect an individual's combinatorial potential. Cupples and Holmes (1992: 270) reach the same conclusion:

> . . . average comprehenders' lexical entries sometimes contain less explicit information about form class than those of good comprehenders. As a result, average comprehenders will sometimes have available to them less information on which to base a syntactic analysis. Consequently, the syntactic representations they construct might be incomplete, forcing them to base their interpretations of sentences on less information about grammatical structure, as such.

They mention results from an unpublished study which showed that individuals with inferior form class knowledge tended to interpret sentences by focusing on the meanings of individual words. This was shown by the fact that such individuals showed large effects of semantic reversibility than individuals with better form class knowledge. In line with the findings of Cupples and Holmes, Chapter 3 reported studies by Chomsky (1969), Kramer Koff and Luria (1972) and Sanders (1971) which suggested that individual differences in lexical knowledge lead to individual differences in the ability to distinguish between sentences containing *ask* or *tell* type verbs.

Chapter 5 also reported work by Perlmutter and MacDonald (1995) which indicates that language users differ in their sensitivity to contextual lexical constraints during the processing of ambiguous sentences. Perlmutter and MacDonald explain these differences in terms of individual differences in sensitivity to lexical probabilities, arising from individual differences in linguistic experience. These differences can therefore be subsumed under the law of frequency. This law should also subsume knowledge of idiomatic expressions, such as greetings, clichés, figures of speech, proverbs, famous sayings, literary quotations and so on. Within this category one might also include recurrent multi-word units such as 'in view of . . .', 'as a matter of fact . . .', 'it is well-known that . . .', 'there is no reason to believe that . . .', '. . . is not what it's cracked up to be' and so on. It is an empirical question just what proportion of written and spoken language such fixed expressions comprise, but it may be considerable. There does not appear to have been an effort to find if native speakers of a language differ in knowledge of idiomatic expressions. Yet clearly they must, given that many fixed expressions are specific to various domains of discourse, such as, for instance, literary quotations or aphorisms.

8.6 Syntax

In the development of syntax in children, Bates *et al.* (1988) report yet another set of dichotomies. At one extreme, early speech is telegraphic and at the other, speech contains rote-learned chunks. This difference relates to the size of units. Another difference relates to combinatory potential. In some children, novel combinations are relatively frequent while in others frozen forms are more frequent. This pattern of differences is related to that discovered by Gleitman and Gleitman (1970) and Geer *et al.* (1971) who found that university students were better at paraphrasing novel compounds such as 'bird house thin' (that is, thin like a bird house) than high school-only educated subjects. To test the possibility that differences in memory capacity might explain the differences in comprehension, Geer *et al.* asked subjects to recall the compounds after each paraphrase. No relationship was found between paraphrase accuracy and recall. To test the possibility that poor performance was due to poor semantic productivity rather than syntactic productivity, Gleitman and Gleitman (1970) used a forced-choice test in which both correct and incorrect alternatives were provided and still obtained effects of education.

In the 1970 study, Gleitman and Gleitman had used three-word compounds. Fillenbaum (1971) proposed that poor performance might be the result of subjects' inability to apply rules recursively, even if they knew the rules. However, when Geer *et al.* (1971) used two-word compounds which involved no recursion, for instance, 'boot green' (a boot kind of green) they still obtained effects of education. The two experiments showed that the high school-only subjects could understand compounds with which they were familiar. Thus, while they could tell the difference between 'dog house' and 'house dog', they could not understand the difference between 'boot green' and 'green boot'.

It therefore appears that the less educated speakers possessed a stock of noun compounds that they had learned by rote and they could not understand novel compounds by rule. The more educated speakers appeared to be aware of the fact that, in English, the syntactic position of a word can determine its lexical class. This inference is based on a more abstract approach to lexical categorisation whereby the lexical class of a word is not associated exclusively with its phonological form but with its syntactic position.

There is evidence that native speakers differ in knowledge of formulaic language. This type of language includes both common syntactic forms that are variously referred to as combinatory types, canonical sentoids, sentence templates or syntactic frames. Townsend and Bever (2001) discuss experimental evidence which suggests that English speakers initially interpret sentences in terms of a fixed A(gent) V(erb) P(atient) schema. Speakers may subsequently alter this interpretation if it is found not to be appropriate. For instance, the temporary ambiguity of sentences like 'The horse raced past the barn fell' can be attributed to the fact that subjects initially interpret 'raced' as the main verb and have to revise this interpretation when additional information comes to light. Mills and Hemsley studied knowledge of permutations of the canonical form (S)ubject (V)erb (O)bject (Q)ualifier. Previous studies by Scott and Mills (1973) had shown that various permutations of this form elicited differing levels of acceptability. Mills and Hemsley's aim in carrying out the study is expressed in the following quotation:

> Demonstrations that certain groups of native speakers of English do not know of the existence of some grammatical forms threatens the validity of [...] [Chomsky's] generalisations, since such ignorance could imply varying levels of competence between individuals. Furthermore, if it were possible to demonstrate that degree of competence co-varied with level of education, it might be possible to suggest that speakers of English progressively learn its grammar, rather than that a complex set of intuitions progressively unfolds. [...] The complex knowledge of highly educated native speakers could then be explained by their greater exposure to and memory for complex linguistic forms (327).

Mills and Hemsley (1976) compared the grammaticality judgement accuracy of university students; individuals with three or four years at high school and individuals with one or two years at high school. Subjects were asked to make grammaticality–acceptability judgements on various permutations of the canonical form: (S)ubject (V)erb (O)bject (Q)ualifier. Mills and Hemsley found that more highly educated subjects were significantly more willing to accept certain permutations as grammatical than the less educated subjects. Similar effects of education in the making of grammaticality judgements are reported by Karanth *et al.* (1996). They found that school-going children (from six to eleven years of age) consistently outperformed non-school-going children of the same age in tasks involving grammaticality judgements and comprehension. A similar pattern of results was obtained by Karanth *et al.* with schooled and unschooled adults. These differences could be described in terms of the frequency with which given forms appear in an individual's linguistic environment. Other examples of frequency effects are described below.

Spencer (1973) compared the degree of agreement in grammaticality–acceptability judgements by professional linguists and students. Using sentences from six linguistics journal articles, Spencer found that students agreed with each other's grammaticality–acceptability judgements 80 per cent of the time, but that they agreed with only 50 per cent of the linguists' judgements. Spencer (1973: 95) suggested that the difference in judgements arose from the frequency with which linguists are exposed to the types of sentences discussed in linguistics journals:

> Linguists' exemplars are said over and over, repeated in various articles and presented at various conferences. Always, these exemplars are used in conjunction with proposed rules. Indeed, exemplars become so well known among linguistic circles that, for instance, the whole controversy involving deep and surface structure can be identified by a linguist saying "eager-easy". Through repetition of exemplar sentences in conjunction with rules and prolonged pondering and inspection of the exemplar cases, the intuitions of acceptability of those cases changes.

Baruzzi (1983) studied the effects of education on the comprehension of a large number of common sentence structures in Italian. She found that more educated subjects performed better than less educated subjects in the Token Test and the Syntax Battery. These tests involve various common syntactic and semantic operations involving modification, assignment of thematic roles, determining the scope of quantification and resolution of anaphora. The results indicate that more educated individuals have mastered a wider variety of syntactic forms than less educated individuals.

Sampson (2000) carried out a syntactic annotation of parts of the British National Corpus. When he analysed syntactic complexity by age group, he found that the level of syntactic complexity rose throughout the lifespan and that 80-year-olds displayed more complex syntactic structures than younger people. This result is counter to the idea that syntactic complexity declines with age (Cheung and Kemper, 1992) on account of declining working-memory capacity. Sampson notes that it is possible that the apparent increase in syntactic complexity is an effect of language change, and that levels of syntactic complexity are actually declining in British English. While there is nothing in the data to discount this possibility, it would be odd, from a rule-based point of view, to suggest that succeeding generations of British English speakers apply syntactic rules less and less recursively and that this is a kind of language change. To do so would be to suggest that grammars set specific limits on the degree of structural complexity. It is, of course, possible that complex syntax is related to formal uses of language. If there is a tendency to use language less and less formally then there would be an accompanying tendency to use less complex syntax.

There is another possibility, however. In Saussure's theory, expressions generated by the language system become part of the system. It is therefore theoretically possible for structures stored in memory to grow incrementally with use. That is to say, previous structures may support the generation of slightly more complex structures, so that more linguistically experienced individuals find it easier to produce structurally complex expressions, particularly if those expressions are overlearned and express content which is highly familiar. This account is, of course, purely speculative and more work is needed in order to interpret Sampson's results.

8.7 Complex syntax

Dabrowska (1997) carried out a study to find out how well the sorts of structures often discussed in linguistics journals can be understood by native English speakers with different levels of education. She used tough movement sentences, for instance, 'Sandy will be easy to get the president to vote for'; complex NP sentences, for instance, 'The manager knew that the fact that taking good care of herself was essential upset Alice', and parasitic gap sentences, for instance, 'It was King Luis who the general convinced that this slave might speak to'. Dabrowska found that university lecturers understood such sentences better than undergraduates who, in turn, understood them better than porters and cleaners. A possible account of these differences in terms of Saussure's theory is suggested below.

In the case of the tough movement sentence, 'Sandy will be easy to get the president to vote for', subjects were asked, 'Who will find it easy to do something?'. The answer is 'someone not mentioned in the sentence' such as the speaker, for instance. Subjects who answered incorrectly suggested that it might be Sandy or the president who would find it difficult to do something. The sentence is derived from a simpler sentence such as 'Sandy will be easy to see'. It would seem obvious from the simpler sentence that the person who will find it easy to do something cannot be Sandy. It should be noted, however, that

Carol Chomsky found that some children interpreted sentences like the simpler tough movement sentence to mean that it would be Sandy who would find it easy to see (though her study, as noted earlier, was criticised on methodological grounds by Kessel, 1970). It should not therefore be automatically assumed that all adult native speakers of English can actually parse the simpler version correctly.

Assuming, for the sake of argument, that all normal adult native speakers of English can parse the simpler tough movement sentence correctly, the question arises as to why they should fail to understand the more elaborate version (leaving aside the obvious working-memory account). A purely syntactic analysis of the sentence would suggest that the object of the infinitive in 'will be easy to see' is the same as that in 'will be easy to get the president to vote for'. For some reason, embedding the infinitive within another makes it impossible for some native speakers to understand the sentence correctly. A possible explanation is that these native speakers may not carry out detailed syntactic analysis (as suggested by Stolz, 1976). It could be that they possess a syntactic formula of the form 'NP Auxiliary Verb Easy To Verb' which they have learned from experience to interpret appropriately. Any marked change to this formula could result in confusion and defaulting to guesswork.

A similar account could be suggested for the parasitic gap sentence, 'It was King Luis who the general convinced that this slave might speak to'. It would seem obvious, yet again, that normal adult native speakers of English can understand a sentence like, 'It was King Louis who the general convinced'. As suggested earlier, it would be unwise to make such an assumption without empirical justification. Assuming, once again for the sake of argument, that the second sentence can be understood by all normal adult native English speakers, it is not clear why elaborating the verb phrase should lead to a comprehension error. A purely syntactic analysis would suggest that the complement of 'convinced' in the simpler sentence is the same as the complement of 'convinced that this slave might speak to' in the more elaborate version. It could be the case, yet again, that subjects who failed to comprehend the more complex sentence did not carry out a detailed syntactic analysis but understood the simpler sentence by using a learned syntactic formula of the form: 'It was X who Verb Y'. The inclusion of additional material in the elaborate sentence which could not then be inserted into this formula would then have resulted in confusion.

The next chapter will present experimental evidence which suggests that failure to understand the complex NP sentence, 'The manager knew that the fact that taking good care of herself was essential upset

Alice' can also be attributed to the imposition of a syntactic formula. This evidence was only obtained incidentally and should therefore be treated as suggestive, given that experiment did not deliberately set out to investigate the use of syntactic formulae. Townsend and Bever (2001) review some experimental evidence which supports the use of formulae, though, yet again, this evidence is only suggestive. The use of formulae is, however, consistent with the theoretical account being discussed in this chapter.

Presumably, subjects who understood the test sentences correctly did so by carrying out more detailed syntactic analysis. This analysis would have involved generating a representation describable in terms of phrase structure, even if the subjects would not have used explicit syntactic symbols. As suggested in the discussion of rewrite rules in the previous chapter, the ability to impose phrase structure might involve the ability to recognise abstract paradigmatic relationships. This account needs further development. If it is valid, however, it would suggest that the more educated subjects possess more abstract syntactic categories than the less educated subjects. This suggestion is similar to one made regarding the subjects in the studies by Gleitman and Gleitman (1970) and Geer *et al.* (1971). In that case, greater exposure to complex structures via academic discourse could have led to the induction of more abstract categories. It is possible, however, that more educated subjects might have understood complex sentences because they possessed more complex formulae. This possibility is suggested in the following paragraphs.

8.8 Self-embedded sentences

Chapter 3 has already described individual differences in the ability to understand self-embedded sentences. Chapter 7 suggested that some individuals might treat sentences with a single embedded clause, such as, 'The man I have seen . . . ' as instances of a syntactic formula of the form, 'N N V V' (leaving aside for the sake of convenience other material associated with a noun phrase, such as determiners, adjectives and so on). Such individuals would then not be able to generalise this formula to sentences with two levels of self-embedding. Support for this interpretation comes from the discussion of Freedle and Craun's experiment in Chapter 3, which showed that, given examples of sentences with two levels of self-embedding, some individuals can only generalise to sentences with two levels of self-embedding but not to sentences with three levels of self-embedding. This behaviour suggests that they had acquired a more complex formula of the form 'N N N V V V'.

The question arises as to what it takes to be able to grasp the notion of arbitrary self-embedding. It is interesting to note that linguists use metaphorical terms in describing the structure of self-embedded sentences. The terms 'nested structure', or 'mirror structure', or even the term 'embedding' itself, invoke spatial patterns from visual or tactile sensory domains. It may be that even linguists find it easier to grasp the structure of self-embedded sentences by transferring their knowledge of spatial relationships to relationships between clauses. Whatever the merits of this suggestion, it appears that the ability to induce rules is related to the ability to transfer patterns across contexts, that is, the ability to generalise, to perceive that which diverse phenomena have in common. This is the principle of similarity in operation, as suggested below.

In purely syntactic terms, the ability to grasp the notion of arbitrary self-embedding requires the recognition that all nouns can be post-modified by a clause, whether the noun in question is in the matrix or embedded clause. This realisation therefore depends on a context-free definition of the noun phrase. It appears that some individuals have difficulty in making the realisation because they seek to represent sentences in terms of global formulae, such as 'N N V V' or 'N N N V V V'. These global formulae are only useful for specific instances. A more general approach requires an appropriately sized and appropriately defined unit, that is, an NP that has the option of being post-modified by an S, regardless of the syntactic context of the NP. The possession of such a unit opens the way for an arbitrary degree of self-embedding. Of course, understanding the *structure* of self-embedded sentences does not entail the ability to understand multiply self-embedded sentences. There is always the possibility of interference between the nouns and between the verbs. It is notable that self-embedded sentences are easier if the noun concepts are distinct from each other (Hudson, 1996). It may therefore be that once the structure of self-embedded sentences is understood, the problem then becomes one of counteracting the effects of interference. In that case, it is necessary to apply the law of contrast in encoding the nouns and verbs in order to ensure that they are distinct from each other and can be reliably retrieved from LTM.

8.9 Text

Work by Gernsbacher and Faust (1991) and Gernsbacher *et al.* (1990) suggests that less skilled readers tend to create disconnected textual representations compared to skilled readers. For instance, skilled and less skilled readers were tested on memory for the surface form of normal and scrambled text. Gernsbacher *et al.* discovered that skilled readers show higher memory for the surface form of normal texts than for scrambled text. Less skilled readers, on the other hand, show the same levels of recall for normal and scrambled text. This suggests that the representations that less skilled readers create for normal texts are similar to those that they create for scrambled text, that is, they are localised and disconnected. Therefore, individuals also differ in the size of textual units.

A similar finding has been reported in studies of individual differences in sentence memory. Anderson (1974b) asked subjects to memorise a series of sentences. The subject of each sentence was in a specified location, for example, 'The hippie is in the park', 'The lawyer is in the park' and 'The hippie is in the bank'. Anderson found that the time needed to recognise each sentence correctly was related to the number of times that its subject or location occurred in the other sentences. This is called a fan effect. Ericsson and Kintsch note, however, that the fan effect is reduced if the series of sentences are thematically related in such a way that they can be integrated into a narrative, as suggested by the following study.

Cantor and Engle (1993) found individual differences in the fan effect related to thematically related and unrelated sentences. They asked subjects classed as having a high or a low working memory to carry out a recognition judgement task on thematically related and unrelated sentences. For the unrelated sentences, both groups of subjects produced a fan effect, though this was higher for the low working-memory subjects. For thematically related sentences, however, low working-memory subjects produced a fan effect but the high working-memory subjects produced a negative fan, that is, recognition judgements were faster the greater the number of times sentence elements occurred across sentences. This result suggests that the high working-memory subjects were integrating the thematically related sentences, perhaps into a narrative. Ericsson and Kintsch (1995: 228) conclude that:

> These findings show simply that the presented sentences are encoded in LTM [long term memory] differently for the two groups. The low-WM group appears to encode the sentences in isolation or in small groups of thematically related ones, whereas the high-WM group is able to attain a more integrated representation of the sentences, especially for the thematically-related sentences.

Skilled readers therefore distinguish similar sentences from each other by stringing them in a narrative. Thus we have the laws of contiguity and contrast in operation, with sequential relationships being used to create contrasts.

8.10 Conclusion

The discussion has shown that Saussure's theory can be used to describe patterns of individual differences reported in the literature. A generalised account of the application of the theory to individual differences is provided below. The account highlights the manner in which individual differences in the application of one law can have an impact on the application of the other laws.

Variations in applying the law of contrast can be expected to lead to variations in the linguistic distinctions made by individuals. Such variations could have a knock-on effect on other aspects of language. They could impact, for instance, on the paradigms or categories created by that individual. That, in turn, would affect the individual's combinatory potential. Further, the distinctions made by an individual could have an impact on their ability to notice certain features in the input, thus affecting their sensitivity to frequency with which those features occur in certain contexts. Variations in the law of similarity would affect the linguistic categories created by an individual in terms of the number of those categories and the level of abstractness which they capture. Differences in categorisation can be expected to affect combinatorial potential.

Variations in applying the law of contiguity would affect the size of units created by an individual. It is interesting to note that all four laws are involved in determining the manner in which units are defined. As stated above, sequential associations come under the law of contiguity. If those sequential associations are highly frequent, then the sequence becomes unitised and can be treated as a single unit, in a manner analogous to block coding in information theory. Unitisation diminishes the contrasts between the sub-units of a sequence, again, in a manner analogous to the fact that block coding masks the internal structure of a sequence. However, if sub-units of a sequence occur frequently in a number of different sequences, they become salient and give rise to a recognition that the 'same' unit occurs in different contexts (the law of similarity). When that happens, the contrasts between sub-units are sharpened and the internal structure of the parent unit becomes more clearly defined. It should be noted that it is not simply the case that smaller units are better than larger units. It is a question of what kind of unit is most efficient given task-specific characteristics of the input.

The next chapter will present experimental evidence which suggests that individuals employ different kinds of syntactic unit. Some individuals appear to impose a syntactic formula on the test sentences in a top-down manner without carrying out a detailed syntactic analysis. Other individuals appear to carry out a detailed analysis and to generate a syntactic representation in a bottom-up manner. Individual differences in unit size are therefore related to the degree to which individuals engage in either bottom-up or top-down processing. This pattern of individual variation appears to be highly general, as it is evident from phonological to text levels of representation. It is reminiscent of the expert–novice differences discussed in Chapter 1. Further, individuals who engage in top-down processing behave in a manner which in some ways is consistent with the pattern-matching processes modelled by the experience-based approach. On the other hand, the performance of individuals who engage in bottom-up processing is more consistent with rule-governed behaviour.

9
Effects of Recall Training and Comprehension Training

9.1 Introduction

The previous chapter presented evidence which indicates that native speakers vary in the manner in which they define their linguistic units. This evidence supports the idea that native speakers differ in the manner in which they represent sentences. If that is the case, then it may be argued that native speaker variations in understanding complex sentences are due to individual differences in syntactic representation rather than inherent individual differences in working-memory capacity (as proposed by Just and Carpenter, 1992). This argument is based on Ericsson and Kintsch's (1995) theory of skilled processing, which states that total working-memory capacity reflects the efficacy of the representations used to encode inputs in long-term memory (LTM). This chapter describes an experiment carried out to find out if native speaker variations in understanding complex sentences are due to variations in working-memory capacity or to variations in syntactic representation.

Before describing the experiment itself, the chapter will first give a brief description of a preliminary experiment. This description provides a necessary background to the main experiment. In addition, the preliminary experiment produced interesting results, only some of which could be followed up in the main experiment described in this chapter. It is hoped that other researchers will seek to replicate some of these results, whose interpretation has a significant bearing on the kinds of representations employed by native English speakers.

9.2 A preliminary experiment

The preliminary experiment sought to determine if there is a relationship between an individual's level of education and their ability to comprehend complex sentences. A number of studies reviewed in previous chapters indicate that there is a strong relationship between syntactic skill and the number of years spent in formal education (Baruzzi, 1982; Dabrowska, 1997; Geer *et al.*, 1971; Gleitman and Gleitman, 1970; Karanth *et al.*, 1996; Mills and Hemsley, 1976). The study by Dabrowska, in particular, indicates that syntactic skill increases steadily with increasing levels of formal education. The relationship between education and syntactic skill suggested to Dabrowska that highly educated *non-native* speakers of English might be more syntactically skilled than less educated native English speakers. The preliminary experiment described below was carried out at her suggestion.

9.2.1 Subjects

In selecting the subjects for the experiment, it was desirable to compare subjects with widely differing degrees of education, as this comparison would make it easier to detect effects of education. It was therefore decided to compare graduate university students against subjects with a high school-only level of education. Graduate native speakers were compared against high school-only native speakers in order to measure the relationship between education and syntactic competence among native speakers. These two groups of native speakers were also compared against highly educated non-native speakers of English in order to measure the relationship between nativeness and education with regard to syntactic competence in English.

The choice of subjects was also based on another consideration. In non-linguistic domains of skill, the most dramatic effects of individual differences have been obtained from comparisons of experts and novices. For instance, there is evidence that child chess experts have higher working memory for chess than adult chess novices (Ericsson and Kintsch, 1995). This result is particularly striking because children are supposed to have lower working-memory capacities than adults. Part of the motivation in carrying out the experiment was to see if such patterns of expert–novice differences can be generalised to syntactic processing. Therefore, the native and non-native graduates who were selected for the experiment were all students in linguistics. The students could be considered as experts in syntactic analysis and the high school-only subjects as novices. This choice of subjects, meant, of course, that the results could not be generalised to speakers with no background in linguistics.

9.2.2 Materials

The test sentences used in the experiment needed to be complex and grammatically novel. The reason for this is as follows. If syntactic knowledge is rule-based, then one would expect native speakers to be able to comprehend such structures successfully. It is important to appreciate that the rule-based approach makes an extremely bold prediction about native speakers in this respect, namely, that they can understand *any* sentence, regardless of its structural complexity or rarity, as long as that sentence can be assigned a structural description by the grammar and as long as performance factors are controlled for. An effective way of testing this prediction is to use complex and unfamiliar syntactic forms. If familiar sentence forms are used, there is danger that subjects might comprehend them using syntactic formulae (see Townsend and Bever, 2001). The description of the materials below focuses on just one of the structures used in the preliminary experiment. This structure was also used in the main experiment, to be described presently; therefore, a detailed discussion of the issues at this point provides a useful background to that experiment.

9.2.3 Some objections

A number of commentators on the preliminary experiment have objected to the use of complex and unfamiliar sentence forms. These objections have often taken two forms. One objection is that the sentences used in the experiment do not occur frequently in natural language and therefore the inability to understand them does not say anything about the normal use of language. Another objection is that the sentences may be ungrammatical for certain groups of native speakers, but not for others. The following paragraphs will address these concerns.

Firstly, the aim of the experiment was, in part, to find out whether knowledge of language really takes the form of rules, rather than to find out about the normal use of language. If knowledge of language does take the form of rules, then native speakers ought to understand sentences constructed on the basis of those rules, regardless of complexity. The experiment would have to ensure that any observed individual differences were not attributable to performance factors. To ensure that the test sentences were grammatical, several steps were taken. Firstly, simple sentences were elaborated recursively, thus ensuring that the resulting sentences were based on a well-established rule. Secondly, the sentences were submitted to two experts in English syntax, who found no grammatical violations in the sentences. Thirdly, corpus queries were carried out on the British National Corpus in order to ensure that the sentence forms used in the experiment did, in fact, occur in natural language.

The sentence structure in question was constructed as follows. Firstly, the verb complement sentence:

> *Tom knows that flying planes low is very dangerous.*

was elaborated recursively by embedding the verb complement clause 'flying planes low is very dangerous' in a complex NP, hence:

> *Tom knows that the fact that [flying planes low is very dangerous] excites the pilot.*

The two expert judges considered the sentence to be a perfectly grammatical sentence of English. Although this form is relatively rare (see Granath, 2001) there are many examples of it to be found in the British National Corpus. Many of these examples are far more complex compared with the test sentences. Some examples are listed below:

> *Bourne, creator of the original Unix Bourne shell command interpreter, believes that the fact that development of Unix operating systems is generally done in a handcrafted manner is not a function of the inherent diversity of Unix itself, but is a long-term software engineering problem.*

> *But one may also say, from the fact that this has needed to be discussed – and from the fact that, as I say, throughout the greater part of Christendom women have not been ordained – it would be difficult to argue that the fact that this religion has had at its centre a male figure has been of little significance.*

> *Consequently, the answer to be given to the national court must be that the fact that the competent minister of a member state has the power to dispense with the nationality requirement in respect of an individual in view of the length of time such individual has resided in that member state and has been involved in the fishing industry of that member state cannot justify, in regard to Community law, the rule under which registration of a fishing vessel is subject to a nationality requirement and a requirement as to residence and domicile.*

These examples show that native speakers of English can and do construct highly complex sentences of the sort used in the experiment, albeit in specialised domains of discourse. The argument that the sentence used in the experiment is ungrammatical therefore appears to have no basis in actual usage.

9.2.4 Controls on working memory

The experiment had to ensure that any individual differences in understanding the test sentences were not due to performance limitations. According to Just and Carpenter (1992), individuals differ in working-memory capacity. Therefore, according to that theory, individual differences in the ability to understand test sentences can be attributed to individual differences in working-memory capacity, rather than individual differences in the ability to represent such sentences. In order to control for differences in working-memory capacity, subjects were allowed to study the sentences visually or as long as they wished. Blaubergs and Braine (1976) showed that subjects could understand sentences with up to five levels of self-embedding if the sentences were presented visually. To illustrate the effect of visual presentation and relaxed time constraints, consider the following example from the British National Corpus:

> *I think it is only right to comment that the fact that it is, in a case such as the present, open to a taxpayer to stipulate, if he wishes, that the money shall be repaid if it is found not to be due in pending proceedings, provides another practical reason why a case such as the present is likely to occur only in rare circumstances indeed.*

It is easy to attribute the difficulty in processing this sentence to limits in STM or working memory. However, the sentence is easily parsable, given sufficient study time and sufficient syntactic knowledge. There are certain syntactic cues in the sentence which enable it to be segmented in a manner which helps to work out the grammatical relations between the segments.

For instance, all material subsequent to 'It is only right to comment that . . . ' comprises the complement clause of the verb 'comment'. All material subsequent to 'the fact that' and ending in 'proceedings' comprises the complement clause of the noun phrase headed by 'fact'. It is then clear that the verb 'provides' is the predicate of the subject noun 'fact'. This analysis, which need not make explicit use of technical grammatical terms, makes it possible to answer comprehension questions such as 'What comment is it right to make?' and 'What provides another reason why "a case such as the present is likely to occur only in rare circumstances indeed?"' and so on. It is therefore implausible to suggest that failure to comprehend the sentence is due to insufficient working-memory capacity. The most likely cause of incomprehension is an inability to exploit the syntactic structure of the sentence in order to work out the grammatical relations which it encodes.

In the experiment itself, subjects had to read sentences such as the following and then answer the accompanying comprehension questions.

> *The doctor knows that the fact that taking good care of himself is essential surprises Tom.*

What does the doctor know?

> *The doctor knows that the fact that taking good care of himself is essential surprises Tom.*

What surprises Tom?

> *That the doctor knows that taking good care of himself is essential.*

9.2.5 Comprehension response types

Subjects' responses were analysed and placed into various categories. Correct responses were often characterised by rearrangement of the order of syntactic constituents in the structure. Incorrect responses, however, were often characterised by a retention of the linear order accompanied by an omission of one or more of the syntactic constituents. These modifications of the structure are described in more detail below.

Rearrangement often involved shifting the complex NP to the end of the sentence. Thus *The doctor knows that the fact that taking good care of himself is essential surprises Tom* became *The doctor knows that Tom is surprised by the fact that it is essential to take good care of himself.* Some subjects compressed the entire complex NP into a much shorter phrase, for instance: *The doctor knows that* ***this fact*** *surprises Tom.*

Omission took several forms, but generally it involved simplifying the test structure in such a way as to make the subject of the verbal complement a simple rather than a complex noun phrase. For instance, some subjects omitted the predicate of the subject of the most embedded complement clause to produce the following type of response: *The doctor knows that [the fact that] taking good care of himself [is essential] surprises Tom* (that is, they missed out the words in square brackets). Other subjects omitted the predicate of the top level complement clause and produced answers such as: *The doctor knows that [the fact that] taking good care of himself is essential [surprises Tom]*. Some subjects produced ungrammatical responses like: *The doctor knows that the fact that taking care of himself is dangerous* or *The doctor knows that the fact that taking care of himself surprises Tom* or *The doctor knows that the fact that taking care of himself*.

Clearly, subjects who carried out rearrangement knew how to decompose the test structure into its constituents and use those constituents to compose a different but semantically equivalent configuration. Subjects who omitted syntactic constituents, on the other hand, did not seem to appreciate the semantic consequences of omitting one or more words in the sentence. It is notable that subjects who carried out omissions invariably retained the word order of the test construction. This inflexibility, coupled with omission of important constituents, suggests that these subjects might have been imposing a syntactic formula on the test sentence. This formula can be characterised in terms of the following schematic sequence <S(ubject) cognitive Verb *that* S(ubject) Verb> where the second subject is realised by a simple rather than a complex NP, for example, *The doctor knows that Tom is lazy*.

The distribution of the response types across groups is shown in Figure 9.1. It is notable that the non-native graduates produced the most

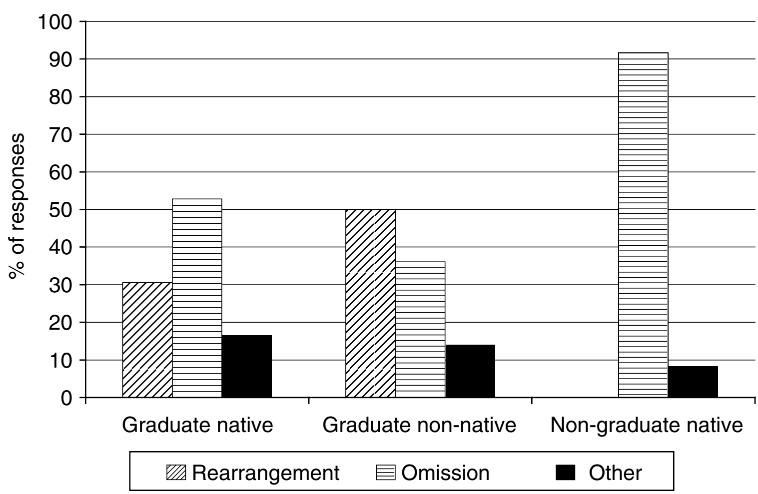

Figure 9.1 Distribution of response types

rearrangements while the non-graduate native speakers made the most omissions. The native graduates performed somewhere in between the other two groups. This result suggests that the non-native graduates carried out more detailed syntactic analysis than both groups of native speakers.

These results mirror those of the whole experiment, including the other structures not described here. There was main effect of groups on comprehension scores for all structures, $F(2,27) = 10.617$, $p < 0.001$. A *post-hoc* Tukey test showed that there was no significant difference between the native and non-native graduates, $p < 0.275$, but the differences between the native graduates and the high school-only native speakers and between the non-native graduates and the high school-only native speakers were significant at $p < 0.001$. All groups performed equally well on control items.

9.2.6 Discussion

This surprising result indicates that certain groups of non-native speakers are more competent in English syntax than certain groups of native speakers. The result was explained by saying that the non-native speakers had obtained explicit instruction in English grammar, both during their learning of English and their study of linguistics. The fact that the non-native speakers tended to rearrange the test sentences in a flexible manner suggests a stronger grasp of syntax, compared to the more rigid approach of the high school-only native speakers. The results also suggested that some individuals may comprehend sentences in a more or less rule-based manner while other subjects do so on the basis of syntactic formulae, which they impose top-down on the input. The fact that foreign languages are often taught using explicit grammatical rules means that experience-based approaches to sentence comprehension may need to acknowledge that, in the case of non-native speakers of a language at least, comprehension can, indeed, proceed in a rule-governed manner.

In general, the results were consistent with well-established expert–novice differences. It appeared that the effect of linguistic expertise was high enough to reverse the normal superiority of native over non-native speakers. The results of the experiment therefore appear to support Ericsson and Kintsch's theory over that of Just and Carpenter. However, the experiment described above had certain limitations. Firstly, the number of subjects was small, ten per group. Secondly, there were no controls on age: the high school-only native speakers were generally older than the graduate subjects. Thirdly, a number of commentators have proposed that working memory might still be responsible for the group differences, despite the use of visual presentation and unlimited study time. A final problem is that the use of subjects with linguistic expertise reduces the generality of the results. Another experiment was therefore carried out to address these limitations. In order to understand the design of the second experiment, it is useful to recapitulate on Just and Carpenter and Ericsson and Kintsch's theories briefly.

9.3 The main study

9.3.1 Rationale

Just and Carpenter (1992) define working memory as the total amount of activation available for symbolic processes. This activation is used to maintain information in an active state (the storage function) as well as to manipulate that information (the processing function). Memory management mechanisms allocate memory dynamically to either processing or storage functions depending on cognitive contingencies. Just and Carpenter propose that there are individual differences in working-memory capacity which manifest in syntactic performance. They retain the assumption that native speakers possess a uniform rule-based syntactic competence, though they do not provide any justification for that assumption.

Ericsson and Kintsch (1995) accept the existence of a limited capacity and domain-independent working memory, which they call short-term working memory (ST-WM). However, they question the idea that individual differences in cognitive processing are constrained by the capacity of ST-WM. Instead, they argue that skilled individuals harness LTM for cognitive processing, effectively creating a long-term working memory (LT-WM). It is in respect of LT-WM capacity that individuals differ. Ericsson and Kintsch cite numerous studies which show that the development of expertise is accompanied by dramatic increases in total working-memory capacity. Ericsson and Kintsch therefore argue that individual differences in comprehension arise from individual differences in linguistic representation rather than from inherent individual differences in a domain-independent ST-WM.

There is a certain difficulty in testing these two theories. Just and Carpenter's theory requires a specific test of working memory – the Daneman and Carpenter working-memory test (Daneman and Carpenter, 1980). The aim of the test is to measure both the processing and storage functions of working memory. In the test, subjects hear a set of sentences and are subsequently asked to recall the last word in each sentence from the set. Subjects are also asked to answer comprehension questions in order to ensure that they carry out both processing and recall operations. The number of sentences in each set increases gradually until recall falls below a certain threshold. The set at which subjects perform above the threshold gives a measure of their working-memory capacity. Just and Carpenter interpret a correlation between working-memory capacity and performance as evidence of an effect of working-memory capacity on comprehension.

However, Ericsson and Kintsch argue that the Daneman and Carpenter test calls upon linguistic skill and therefore favours individuals who are highly skilled in language. They cite studies which show that success in the test is related to the ability to store and retrieve information from LTM in an efficient manner. In their theory, all learned skills involve the ability to make efficient use of LTM to boost working-memory capacity. Therefore, they argue that superior performance in the Daneman and Carpenter test simply reflects superior sentence encoding skills.

A correlation between working-memory scores and comprehension scores can therefore be interpreted in two completely different ways: as an effect of working-memory capacity on sentence encoding operations, or as an effect of sentence encoding operations on recall performance. Consequently, such correlations cannot be used to discriminate between the two theories. The experiment therefore did not seek to measure the working-memory capacities of subjects, given that working memory is an ambiguous construct.

An alternative test of the two theories involves training studies. It has been known at least since Miller (1956) that STM capacity can be boosted through training. If it is possible to boost subjects' STM for test sentences through memory training, then, according to Just and Carpenter's theory, extra working capacity should become available for processing functions. The result should be an improvement in comprehension. On the other hand, while Ericsson and Kintsch's theory also predicts a beneficial effect of memory training on sentence recall, it does not predict any effect on comprehension. Two other predictions can be derived through a similar line of reasoning. Just and Carpenter's theory, as it is described in their (1992) paper, predicts no effect of comprehension training on comprehension because subjects are already perfectly competent. On the other hand, Ericsson and Kintsch's theory attributes poor performance to insufficient syntactic competence. They would therefore predict that comprehension training will improve comprehension. Ericsson and Kintsch's theory also makes a secondary prediction. In their theory, accurate sentence encoding processes create LTM traces which facilitate recall. Therefore, comprehension training ought also to improve sentence recall (although memory training will not improve comprehension). The two theories therefore make opposing and testable predictions about the effects of recall and comprehension training on recall and comprehension. These predictions were tested in the experiments reported below.

It is important to note that Just and Carpenter do discuss possible effects of practice on comprehension. In the (1992) article, effects of training are said to lead to improvements in the efficiency of existing procedures through practice. The CC-reader, a computational model which implements Just and Carpenter's theory, contains a competence grammar with recursive rules in the form of an ATN (augmented transition network). Noun phrases and verb phrases are constructed by procedures which are activated whenever elements of a noun phrase or a verb phrase are encountered in the input. Activation is required for each procedure to execute. If high demands on memory lead to insufficient activation, then a procedure may either fail to execute or execute inefficiently. The manner in which practice can make a procedure more efficient is not made clear. Elsewhere, it has been suggested that after a procedure has been executed, its activation decays slowly. This residual activation will then enable that procedure to be activated more easily at a later point in time. This is the account given by Pickering and Branigan (1998) to explain the mechanism of syntactic priming. Another possibility which Pickering and Branigan consider is that the firing threshold of a procedure is lowered each time it is used, thus making it easier for that procedure to be activated (Branigan *et al.*, 1997).

A problem with both accounts is that clauses, noun phrases and verb phrases are found in every complete sentence. The procedures for constructing these representations therefore ought to exist in a permanent state of peak efficiency. It is therefore not clear why a given procedure should be less efficient when it is applied recursively, as in the case of the complex NP sentences in the preliminary experiment described above, than when it is not applied recursively, as in the case of simple sentences. Thus residual activation or lowering of threshold accounts cannot explain increases in efficiency through practice. It is unclear, in any case, if Just and Carpenter (1992) would subscribe to the accounts proposed above, since they do not discuss the mechanism by which practice increases the efficiency of a procedure. Whatever the merits of the idea that practice increases efficiency, measures were taken to ensure that the kinds of training used in the experiment could not be construed as providing practice and thereby increasing the efficiency of existing procedures.

It should also be noted that effects of recall training are not well described by Just and Carpenter either. The precise effects of recall training on comprehension are not clear because Just and Carpenter do not provide a means of quantifying memory requirements, as do, for instance, Gibson and Thomas (1999) (see Chapter 4). There is a possibility that the amount of memory capacity which is released by recall training might not be sufficient to make any difference to processing. This possibility highlights a central problem that afflicts capacity limitation theories in general. According to Navon (1984), resource requirements can only be determined on the basis of performance limitations. That is to say, if one task is found to be more difficult than another, then the difficult task is said to consume more resources. On the other hand, a task which consumes considerable resources is said to be more difficult. This effectively means that working-memory requirements are determined in a circular fashion. As a result, capacity limitation theories are unfalsifiable unless they present independent measures of the capacity taken up by specific operations. No measures of this sort are presented in Just and Carpenter (1992). Therefore, the best prediction that can be drawn from the theory as it stands is that a reduction in storage demands due to recall training will lead to improvements in processing.

9.3.2 Materials and subjects

The experiment involved the same type of sentences used in the preliminary experiment described above. Subjects were 39 native speakers of English with an average age of 18 years. All were junior college students who had undertaken their GCSE examinations in the previous year. Pilot experiments had shown that the ability to understand complex NP sentences used in the preliminary experiment was related to students' performance in the GCSE examinations. Students who had obtained Grade A in English and in four other subjects were found to reliably understand the sentences. On the other hand, students who had obtained Grade B or less in English and four other subjects were found unable to comprehend the sentences correctly. The precise basis for this pattern of difference was irrelevant to the aims of the experiment and no attempt was made to delve into the matter. The critical issue is that these differences are explained differently by the two theories being tested. Just and Carpenter's theory attributes the differences in comprehension to differences in working-memory capacity. Ericsson and Kintsch's theory attributes the differences in comprehension to differences in sentence representation. Academic achievement in national examinations was used purely as a reliable way of assigning subjects to groups.

9.3.3 Procedures

More specific details about the procedures used are presented in the *Procedures* section of the experimental reports. In general, subjects had to carry out a recall task and a comprehension task. The purpose of using the two tasks was to observe relationships between patterns of comprehension and recall. The use of the two tasks was also necessary in order to observe the separate effects of memory training on comprehension and recall and the effects of comprehension training on comprehension and recall. The comprehension task employed off-line visual presentation in order to eliminate storage demands during comprehension (see Blaubergs and Braine, 1976).

Memory training involved rote memorisation. For a subsidiary interest, chunking test sentences into three-word groups was encouraged in one of two training conditions. The aim of encouraging chunking was to find out if an explicit organisation of test sentences would facilitate memory training. In comprehension training, subjects learned to paraphrase examples of the test structure. A paraphrase task was used to ensure that training was not task-specific. For a subsidiary interest, comprehension training in one condition was accompanied by a graphic representation of the test sentence. The aim of the graphic representation was to find out if an explicit knowledge of the structure of the sentence would facilitate comprehension training.

9.3.4 Organisation

The experiment involved a number of sub-experiments. For organisational convenience, each of these sub-experiments will be treated as a full experiment in its own right, with its own set of predictions, methods and results. Experiment I tested the ability of students of high and low academic achievement (HAA and LAA) to comprehend and recall complex NP sentences. This experiment also served as a pre-test for subsequent experiments. Experiment II provided recall training for LAA subjects. The training was specific to the type of structure being tested and was not intended to result in a general increase in memory span. The experiment then tested the effects of recall training on the comprehension and recall of complex NP sentences. Experiment III provided comprehension training for the LAA subjects who had undertaken Experiment II. The comprehension training was specific to the structure being tested and was not intended to result in a general increase in comprehension skill. The experiment then tested the effects of comprehension training on the comprehension and recall of complex NP sentences.

It became apparent after the results from Experiment III had been analysed that the recall training in Experiment II created a ceiling effect which made it impossible to measure the effects of comprehension training in Experiment III on recall. Therefore, another experiment was undertaken to measure effects of comprehension training on recall. A separate group of LAA subjects were provided with comprehension training. Experiment IV then measured the effects of comprehension training on the comprehension and recall of complex NP sentences.

Each experiment is now described below.

9.3.5 Experiment I

Aim

To compare levels of comprehension and recall of complex NP sentences by subjects of HAA and LAA.

Predictions

On the basis of pilot studies, it was predicted that HAA subjects would obtain higher levels of comprehension and recall than LAA subjects.

Methods

Subjects

Group 1 (LAA) was made up of 18 native speakers of English (mean age = 18 years) who, in the previous year, had failed their GCSE English examination or had obtained poor grades (grade 'D' or below) and were now re-taking the subject. Group 2 (HAA) consisted of 11 native speakers of English (mean age = 18 years) who had obtained 'A' grades in GCSE English and in at least four other subjects during the previous year and were now studying for their 'A' levels. Subjects were recruited from local colleges through the help of their teachers and paid £5.00 per hour for participation.

Materials and design

Thirty complex noun phrase complement sentences were used in all four experiments. Comprehension and recall tasks were carried out on ten test sentences. A list of all the sentences and some example questions and answers are provided in the Appendix.

Procedure

Recall test procedure

In the recall task, subjects read 15-word complex NP sentences from a computer screen. Each sentence was prefaced by a get-ready prompt consisting of three plus signs (+++). Each word appeared in the centre of the screen for one second and was then replaced by the next word. After the 15th word, a recall prompt appeared in the form of a question mark (?). Subjects' verbal recalls were recorded on audio-tape for later transcription. Clicking on a mouse button allowed subjects to progress to the next sentence. Two practice sentences were provided, followed by ten test sentences.

Comprehension test procedure

In the comprehension task, subjects read the same sentences they had encountered in the recall task. Each sentence was completely visible on the screen for as long as subjects wished. Clicking on a mouse button caused the first question to appear below the sentence. After answering this question, subjects clicked on the mouse button to replace the first question with the second question. The sentence remained visible at all times. In order to move onto the next sentence, subjects clicked on the mouse. Comprehension responses were also audio-taped for later transcription.

Results

Two scoring systems were used for the recall task. In the non-verbatim scoring system, 1 mark was given for each word which was either the same as that in the test sentence or belonged to the same grammatical class, for example, noun, verb, intensifier, determiner, complementiser. In the verbatim memory scoring system, a single mark was given only for verbatim recall of each word. Results from each scoring system were analysed separately. In some instances, however, scoring system was treated as a factor in its own right in order to study group differences in precision of recall. For the comprehension task, a single mark was awarded for each correct response. Results from each task were subjected to analysis of variance by subjects and by items (F1 and F2). The HAA group scored higher in comprehension than the LAA group, $F1(1,27) = 44.37$, $p < 0.001$; $F2(1,29) = 199.79$, $p < 0.001$ (Figure 9.2).

There was an interaction between groups and items $F(9,243) = 2.18$, $p < 0.02$ which might indicate a practice effect for the HAA group but not for the LAA group (Figure 9.3).

The difference between groups in overall recall scores was significant, $F1(1,27) = 26.27$, $p < 0.001$; $F2(1,29) = 186.96$, $p < 0.001$ for non-verbatim

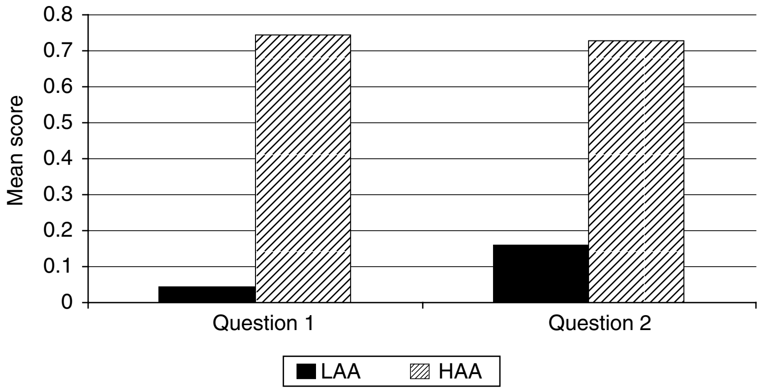

Figure 9.2 HAA and LAA comprehension scores for Experiment I

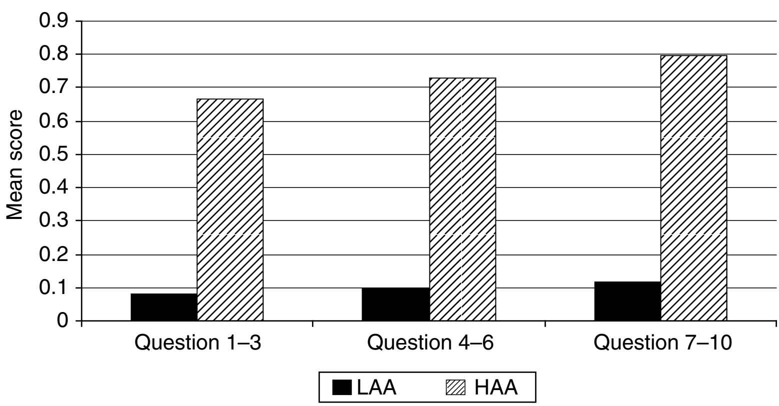

Figure 9.3 Practice effect in Experiment I

memory and for verbatim memory, $F1(1,27) = 28.26$, $p < 0.001$; $F2(1,29) = 158.99$, $p < 0.001$ (Figure 9.4).

The recall curves for the two groups (using results from the non-verbatim scoring system) show two dips, one at the second segment, and another at the fourth segment (Figure 9.5).

The difference between segments is significant for non-verbatim memory, $F1(4,108) = 43.16$, $p < 0.001$; $F2(4,116) = 107.41$, $p < 0.001$ and

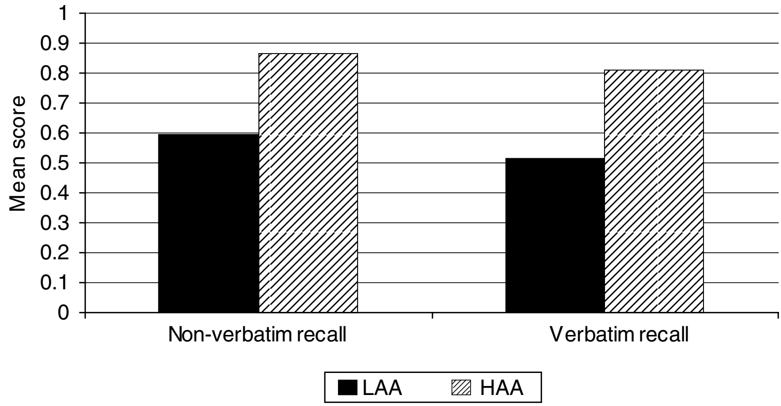

Figure 9.4 HAA and LAA recall scores in Experiment I

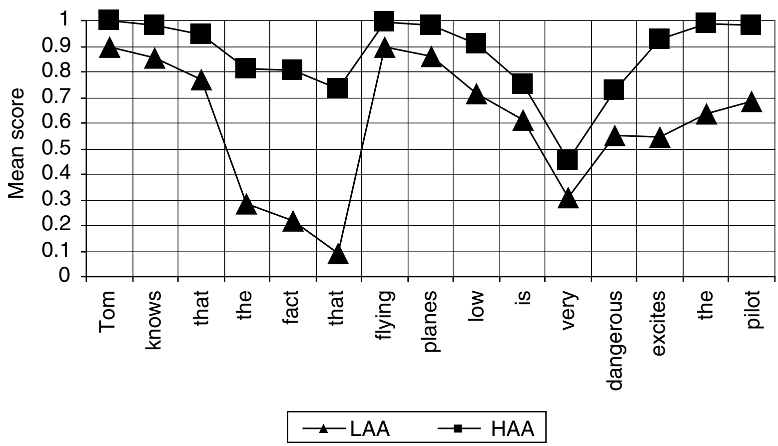

Figure 9.5 HAA and LAA recall curves by word in Experiment I

also for verbatim recall, $F1(4,108) = 31.91$, $p < 0.001$; $F2(4,116) = 84.46$, $p < 0.001$. There were group differences in recall curves, just as there were group differences in comprehension scores: the dips at segments 2 and 4 were much shallower for the HAA group (Figure 9.5). In fact, for this group, the dip in segment 4 was mainly due to poor performance on a single lexical item (an intensifier) which is not relevant to the overall syntactic structure. There was an interaction between groups and segment for both non-verbatim memory, $F1(8,144)=5.187$, $p<0.001$; $F2(4,348)=13.35$, $p<0.001$ and verbatim memory, $F1(8,144)=4.985$, $p<0.001$; $F2(4,348)=9.7$, $p<0.001$.

If recall curves are displayed by segment rather than by word (Figure 9.6), it turns out that the segments which accrued more than 50 per cent correct scores correspond to the simplified version of the test sentences with the form 'Tom knows that flying planes low excites the pilot.' Subjects who failed to comprehend the test sentences often simplified the test sentences in this way. This is also the same pattern of simplification reported in the preliminary experiment described earlier. Comprehension response types from this experiment were not quantified and the interpretation suggested here is based on the quantification of comprehension response types in the preliminary experiment.

There were also group differences in the precision of recall. The HAA group had more precise recall than the LAA group. If scoring system is analysed as a factor in its own right (with two levels – non-verbatim recall and verbatim recall), then there is main effect of scoring system, $F1(1,27)=106.95$, $p<0.001$; $F2(1,29)=107.1$, $p<0.001$ and an interaction between groups and scoring system, $F1(1,27)=9.34$, $p<0.01$; $F2(1,29)=7.36$, $p<0.01$ (see Figure 9.4).

Discussion

The HAA group obtained higher scores in comprehension and recall than the LAA group. These findings are consistent with both Just and

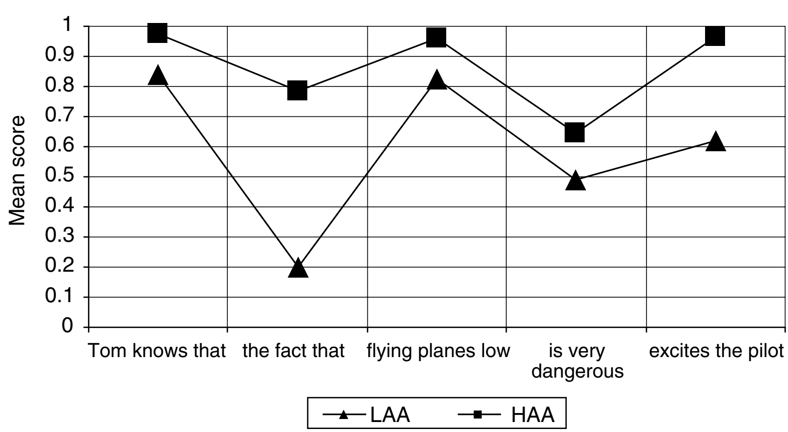

Figure 9.6 Recall curves by segment in Experiment I

Carpenter's theory and Ericsson and Kintsch's. It could be that group differences in comprehension arise from group differences in memory capacity, or it could be that group differences in recall arise from group differences in encoding during comprehension. There appeared to be a relation between comprehension responses and recall curves. This possibility is based on a comparison of comprehension data from the preliminary experiment with recall patterns obtained in this experiment. As in the preliminary experiment, subjects generally interpreted the test sentences in terms of a verb complement structure in which the subject of the complement clause is a simple rather than a complex NP. The recall curves indicate that recall was poorest at just those segments of test sentences which could not be integrated into the simple interpretation produced by the subjects. This pattern of response is most marked for the LAA group.

Finally, there was a possible practice effect for the HAA group which was absent in the LAA group. The performance of the HAA group improved significantly as the experiment progressed. Subjects in this group commented that they gradually became aware that they were misinterpreting the sentences and that when they realised this, they corrected themselves. This behaviour might be explained in terms of a beneficial effect of practice on sentence comprehension procedures, as suggested by Just and Carpenter. The question then would be why there was no such effect for the LAA group. Precision of recall was also higher for the HAA group. This finding is consistent with Ericsson and Kintsch's proposal that skilled comprehenders encode information more accurately than less skilled comprehenders.

9.3.6 Experiment II

Aim

To measure the effect of recall training on the comprehension and recall of complex NP sentences.

Predictions

Just and Carpenter's theory predicts that recall training will boost working-memory capacity and thereby lead to improvements in comprehension and recall. Ericsson and Kintsch's theory predicts no benefit for comprehension though recall is predicted to improve.

Methods

Subjects

The LAA group who participated in Experiment I went to carry out this experiment.

Materials and design

A different set of ten complex NP sentences was drawn from the list of 30 sentences and used in this experiment.

Training procedure

In the training phase, subjects read 15-word sentences. Each word was visible in the centre of the screen for one second and was immediately replaced by the next word. Two methods of memory training were used. For subjects in memory training method A, a one-second pause was inserted after every three words to help subjects group the words into three-word segments. Subjects in memory training method A also received explicit instructions to group the training sentences into three-word segments. In memory training method B, no pauses in sentence presentation or instructions to group were received by subjects. The purpose of using two methods was to see if grouping would facilitate recall training.

After the 15th word had appeared, subjects in both training conditions saw a recall prompt in the form of a question mark (?). Subjects then attempted to recall the sentence aloud to the experimenter. If errors were made, subjects were given feedback on the words they had forgotten and the sentence was presented again. The training cycle was repeated until verbatim recall was achieved. It should be emphasised that the aim of the procedure was to increase subjects' specific ability to recall the test structure and not to bring about a general increase in short-term recall of sentences.

Recall test procedure

In the recall task, subjects read 15-word complex NP sentences from a computer screen. Each sentence was prefaced by a get-ready prompt consisting of three plus signs (+++). Each word appeared in the centre of the screen for one second and was then replaced by the next word. After the 15th word, a recall prompt appeared in the form of a question mark (?). Subjects' verbal recalls were recorded on audio-tape for later transcription. Clicking on a mouse button allowed subjects to progress to the next sentence. Two practice sentences were provided, followed by ten test sentences.

Comprehension test procedure

In the comprehension task, subjects read the same sentences they had encountered in the recall task. Each sentence was completely visible on the screen for as long as subjects wished. Clicking on a mouse button caused the first question to appear below the sentence. After answering this question, subjects clicked on the mouse button to replace the question with the second question. The sentence remained visible at all times. In order to move onto the next sentence, subjects clicked on the mouse. Comprehension responses were also audio-taped for later transcription.

Results

Effects of training on comprehension and recall were measured in terms of the factor phase, which had two levels – pre-test and post-test. There was a significant effect of phase on recall for verbatim memory, $F1(1,17) = 64.75$, $p < 0.001$; $F2(1,29) = 30.8$, $p < 0.001$ and for non-verbatim memory, $F1(1,17) = 26.16$, $p < 0.001$; $F2(1,29) = 15.15$, $p < 0.001$. Figure 9.7 shows pre- and post-test recall scores.

There was no significant difference between pre-test HAA recall and post-test LAA recall for non-verbatim memory, $F1(1,27) = 0.14$, $p < 0.71$; $F2(1,58) = 0.76$, $p < 0.39$. There was, however, a significant group difference in verbatim memory $F1(1,27) = 5.03$, $p < 0.05$; $F2(1,58) = 9.57$, $p < 0.01$ (Figure 9.7). Thus, even after training, the recall of the LAA group was still less precise than that of the HAA group. In fact, although both verbatim and non-verbatim memory increased for the LAA group, there was also an increase in the difference between the two types of recall relative to Experiment I. The interaction of phase and scoring system was not significant by subjects, $F1(1,17) = 3.86$, $p < 0.07$, but it was significant by items, $F2(1,29) = 18.71$, $p < 0.001$.

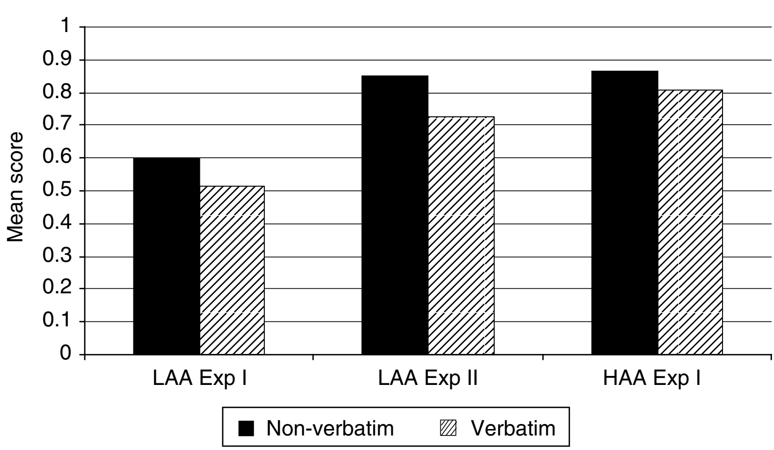

Figure 9.7 LAA and HAA recall scores in Experiments I and II

There was no significant effect of memory training method on recall. More importantly, memory training did not improve comprehension (Figure 9.8). In fact, there was a drop in comprehension scores, though this drop was not significant.

The dips in the recall curve were now shallower than in Experiment I (Figure 9.9), but there is still an effect of segment for both non-verbatim memory, $F1(4,68)=9.83$, $p<0.001$; $F2(4,116)=22.53$, $p<0.001$ and verbatim memory $F1(4,68)=7.32$, $p<0.001$; $F2(4,116)=13.84$, $p<0.001$. Figure 9.9 shows that the recall curves of HAA Experiment I recalls closely resemble that of LAA Experiment II recall. Comparing HAA Experiment I and LAA Experiment II recall data, there was no significant interaction between groups and segment for non-verbatim memory by subjects $F1(4,108)=1.44$, $p<0.225$, but the interaction was marginally significant by items $F2(4,232)=2.93$, $p<0.05$. For verbatim memory, the interaction was not significant by subjects, $F1(4,108)=1.91$, $p<0.11$ but it was significant by items, $F2(4,232)=4.51$, $p<0.002$. The lack of an effect by subjects indicates that recall training enabled the LAA group to perform on par with the HAA group in the recall task.

Discussion

The outcome of this experiment was that recall training facilitated recall but did not affect comprehension. In terms of Just and Carpenter's theory, one would have expected improvements in recall to be accompanied by improvements in comprehension. In that theory, memory is dynamically

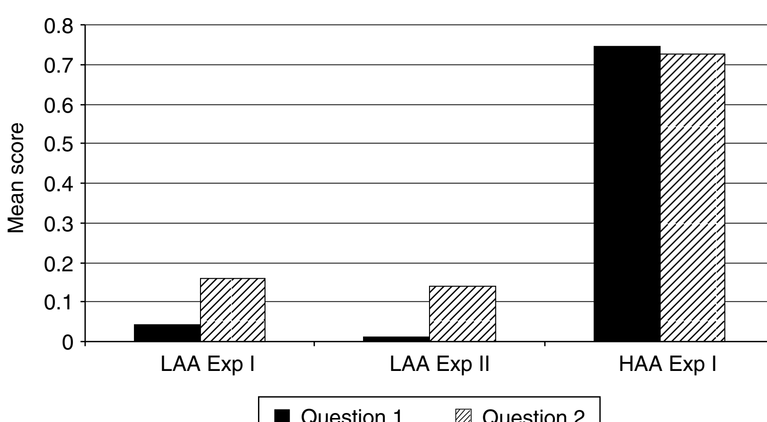

Figure 9.8 LAA and HAA comprehension scores in Experiments I and II

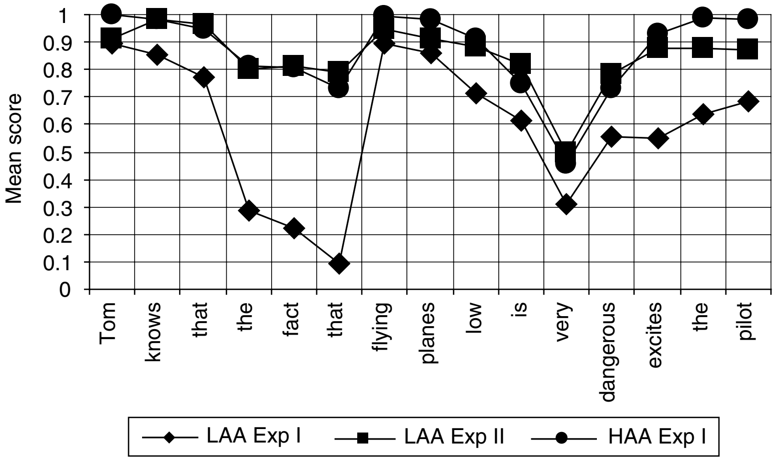

Figure 9.9 HAA and LAA recall curves by word in Experiments I and II

allocated to storage and processing functions. Recall training should therefore have freed up some extra capacity for processing.

It could be suggested that the amount of memory freed up by recall training was below the threshold needed for successful comprehension. This possibility is impossible to test without a metric to quantify memory requirements. Without such a metric there is a danger of circularity, in which task difficulty is a measure of resource requirements and resource requirements are measured by task difficulty (see Navon, 1984 for this critique of resource-based theories). The results are, however, compatible with Ericsson and Kintsch's theory. In that theory, recall can be improved through training. However, the memory representations formed by recall training would have nothing to do with comprehension. Recall training is therefore not predicted by that theory to have an effect on comprehension.

9.3.7 Experiment III

Aim

The aim of the experiment was to measure the effects of comprehension training on comprehension and recall of complex NP sentences. Just and Carpenter's theory predicts no effects, because the theory attributes native speakers with uniform and complete syntactic competence. Ericsson and Kintsch's theory, on the other hand, predicts improvements in both comprehension and recall.

Subjects
The same subjects as those in Experiment II participated for £5.00 an hour.

Methods
Materials
The materials described in the report of Experiment I were used.

Procedure
Two methods of comprehension training were used to evaluate a secondary issue: whether different types of training would lead to different outcomes. In comprehension training method A, subjects were asked to paraphrase training sentences, which were of the same type as those described in Experiment I. Paraphrasing involved two stages. In the first stage, subjects were taught to break the training sentences into two simple sentences which retained the meaning of the training sentence. In the second stage, subjects were taught to recombine the two simple sentences into one sentence which also retained the meaning of the original sentence but had a simpler structure (Table 9.1).

The form of the paraphrase was chosen on the basis of pilot studies which showed that subjects could understand it easily. After completing ten practice items, subjects had to do a further ten items, except that in this case they had to convert a Stage 2 paraphrase into the structure of the training sentence. Training therefore involved 20 items, some of which were also used for demonstration purposes. These items had already been encountered by subjects in Experiments I and II.

In comprehension training method B, subjects were shown a tree diagram representation of a training sentence. This diagram was based on

Table 9.1 Examples of materials for the paraphrase task

Training sentence	Tom knows that the fact that flying planes low is dangerous excites the pilot.
Stage 1 paraphrase	Tom knows that a fact surprises the pilot. The fact that surprises the pilot is that flying planes low is dangerous.
Stage 2 paraphrase	Tom knows that the fact that surprises the pilot is that flying planes low is dangerous.

three-word semantically based units rather than phrase structure. These three-word segments are identical to the ones used in memory training. The purpose of the diagram was to show subjects the relationships between segments of the sentence and the functional roles played by each segment. After the diagram was explained to subjects, various segments were chosen at random from different training sentences (for example, 'excites the pilot') and subjects were asked to identify the node in the tree diagram which represented that segment (in this case, the node labelled 'the effect of something on someone'). When subjects became proficient at labelling the segments, they were given a blank tree diagram and asked to label each node and also fill in the terminal nodes with lexical items from a chosen training sentence. After subjects mastered this task, they were given the same training in paraphrasing as the first group.

The test phase of the experiment was divided into two parts. Pilot studies had shown that subjects did not necessarily transfer their comprehension training to comprehension test. It was therefore decided to prompt subjects to apply their comprehension training halfway through the test phase in order to see if prompting would have a measurable effect. The term Phase in the Results below refers to a comparison between subjects' performance in Experiment II and Experiment III. Note also that, for the comprehension test, Phase involves a comparison between subjects' performance in Experiment II and either the first or second part of the comprehension test in Experiment III.

Results

All subjects learned to paraphrase the training sentences correctly. No systematic data was taken on performance during comprehension training. However, there were notable individual differences in the number of trials needed before fluent paraphrasing was achieved. Some subjects needed only two or three trials while others needed up to five or six trials.

There was no effect of Phase on recall (that is, the effect of comprehension training as measured by the differences in scores between Experiments II and III) for non-verbatim memory $F1(1,17) = 0.56$, $p < 0.466$; $F2(1,29) = 0.05$, $p < 0.824$ or verbatim memory, $F1(1,17) = 2.52$, $p < 0.1305$; $F2(1,29) = 0.688$, $p < 0.4136$. The lack of an effect of comprehension training on recall is attributable to a ceiling effect due to memory training in Experiment II.

There was an effect of Phase on comprehension for the first part of the comprehension test (that is, before subjects were prompted to transfer training to test), $F1(1,17) = 10.479$, $p < 0.01$; $F2(1,4) = 165.77$, $p < 0.001$. There was also an effect of Phase on comprehension for the second part of the comprehension test (that is, after prompting subjects to transfer training to test), $F1(1,17) = 21.21$, $p < 0.001$; $F2(1,4) = 300.5$, $p < 0.001$ (Figure 9.10).

Considering the first part of the comprehension test, there was differential transfer of comprehension training. Seven out of 18 subjects completely failed to transfer and obtained a score of zero. Of the 11 who did transfer, scores ranged from 0.2 to 1.

Summary of effects of comprehension training

There was no significant effect of comprehension training on recall. This is probably due to a ceiling effect on recall resulting from memory training earlier. Comprehension did improve after comprehension training. This finding is consistent with Ericsson and Kintsch's theory. However, the result is difficult to interpret in terms of Just and Carpenter's theory, as discussed earlier. The results also show that there was differential transfer of comprehension training, with slightly less than half of the subjects obtaining zero scores. For those who did transfer, there was variation in scores between individuals, again indicating differential rates of transfer. Prompting subjects to transfer led to a significant rise in comprehension scores. This indicates that, while subjects had the knowledge to comprehend the sentences, they did not always apply it.

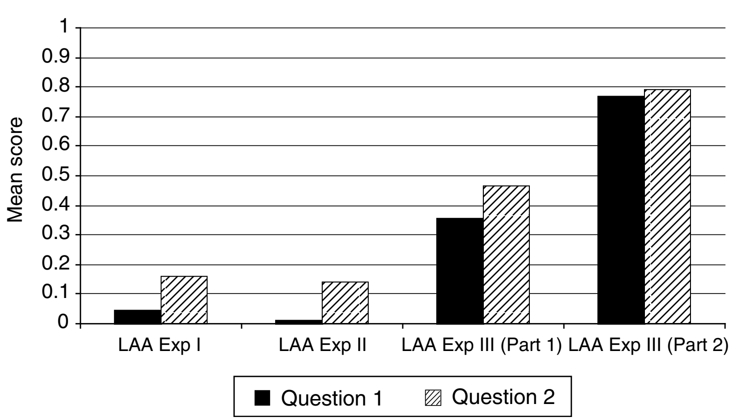

Figure 9.10 LAA comprehension scores in Experiments I, II and III

A new group of subjects was tested to measure the effect of comprehension training on recall in Experiment IV. In view of subjects' inability to automatically transfer training to test, it was decided to use a different training method. The justification for changing the training method is as follows. In the comprehension test task, subjects presumably paraphrased the test sentences, as they had been trained to, in order to form an accurate sentence representation which could then be used to answer the comprehension questions. However, this strategy might not work in the recall task, when subjects had to respond quickly or risk forgetting the words in the sentence. It was therefore decided to train subjects directly on the comprehension task so that they could form an appropriate representation online for use in the recall task. Given that Experiment III had shown that subjects could answer comprehension questions correctly if taught an appropriate strategy, there was no need to repeat this finding in Experiment IV. The main aim of Experiment IV was to find out if comprehension training would have an effect on recall. Note that there would be no task effect, given that the test task (recall) would be different from the training task (comprehension).

9.3.8 Experiment IV

Aim

The aim of this experiment was to investigate the effects of comprehension training on the recall of complex NP sentences.

Predictions

According to Ericsson and Kintsch's theory, comprehension training should improve both comprehension and recall. Just and Carpenter's theory makes no testable predictions about the effects of comprehension training on recall.

Methods

Subjects

Subjects were ten native speakers of English (mean age = 18 years). All except two subjects had obtained less than grade 'A' in the GCSE 'O' level English examinations, the previous year. All subjects were carrying out post-GCSE study. Pilot studies had shown that subjects with less than four 'A' grades at GCSE had difficulties in comprehending the test structure. Thus, although the group used in this study could not all be categorised as having LAA, as those in Experiments I and II, they all fell below the threshold of four 'A' grades. The subjects employed for this experiment will be referred to as the medium academic achievement group, MAA. Comparisons will be presented later to show that this group was comparable to the LAA group in Experiments I and II with respect to comprehension and recall of the test structure.

Materials and design

Materials were the same as those used in Experiment II. The experiment was composed of a pre-test, comprehension training and post-test. In the pre-test, subjects were given two practice items before doing recall and comprehension tasks involving ten sentences. In comprehension training, subjects were given training on the comprehension task (see *Procedure* section for details). Sentences from the pre-test were used for training. In the post-test, subjects were given comprehension and recall tasks involving ten sentences which were different from those used in the pre-test and comprehension training.

Procedure

Pre-test

The procedure is exactly the same as that used in Experiment I.

Comprehension training

During training, subjects were given a non-technical tutorial concerning the test structure, described below. Subjects were first asked to read a training sentence and answer Question 1. In most cases, subjects treated the structure as a simple verb complement structure, which meant that they omitted the second and fourth segments (see Experiment I). Subjects were shown that, in order to be able to integrate the omitted segments, they would have to treat the complex NP as the subject of the complement clause, rather than treating it as the entire complement clause. Subjects were then shown that, in order to treat the complex NP as the subject of the complement clause, they had to take account of the second segment, which provides cues to the effect that the subject of the complement clause is a complex NP.

Subjects were then given an example sentence and answers to both comprehension questions. Subsequently, subjects read practice items and attempted to answer the two comprehension questions. In the event of an error, the explanation described above was repeated and the correct answer to the question was provided. At most, subjects needed three practice items before they could answer the comprehension questions correctly. However, to ensure that subjects had mastered the training, a total of ten practice sentences were given.

Post-test

The test phase consisted of a recall task consisting of ten sentences which was followed by a comprehension task consisting of the same ten sentences. The procedure was exactly the same as that used for the pre-test.

Results

Results were analysed in the same way as those in Experiment I. Where appropriate, the results were analysed in comparison with those of the HAA and LAA groups in Experiment I. The effect of training was analysed in terms of the factor Phase, which had two levels – pre-test and post-test.

Phase 1 (pre-test)

The results mirror those found in Experiment I. The MAA group obtained much lower scores than the HAA group. The group difference in comprehension scores was significant, $F1(1,19) = 9.67$, $p < 0.01$; $F2(1,29) = 197.15$, $p < 0.001$ (Figure 9.11). The MAA group also scored significantly less than the HAA group in both non-verbatim memory, $F1(1,19) = 12.54$, $p < 0.01$; $F2(1,29) = 90.76$, $p < 0.001$ and verbatim memory, $F1(1,19) = 15.17$, $p < 0.001$; $F2(1,29) = 130.95$, $p < 0.001$ (Figure 9.12).

There were no significant differences between the MAA group and the LAA group in terms of comprehension by subjects, $F1(1,26) = 3.013$,

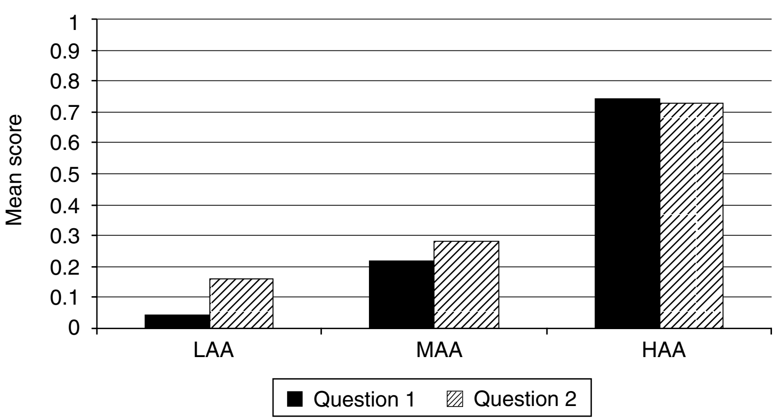

Figure 9.11 LAA, MAA and HAA comprehension scores

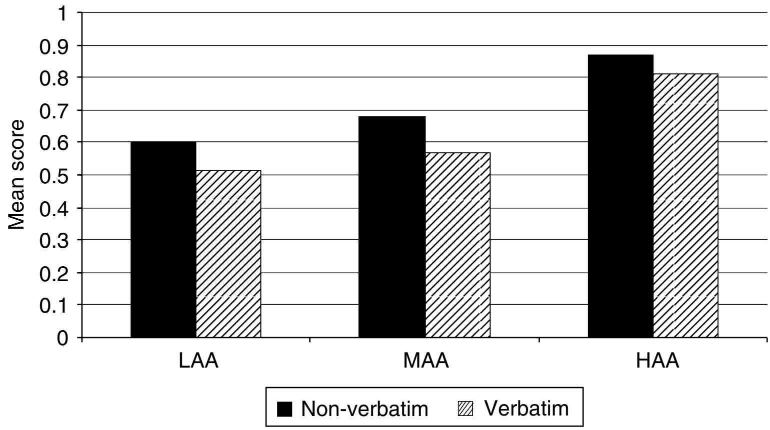

Figure 9.12 LAA, MAA and HAA recall scores

$p < 0.994$ (Figure 9.11); verbatim recall by subjects $F1(1,26) = 0.956$, $p < 0.337$ or verbatim recall by items, $F2(1,58) = 2.628$, $p < 0.111$ (Figure 9.12). However, there were significant differences by items for both comprehension, $F2(1,29) = 86.751$, $p < 0.001$ and non-verbatim recall, $F2(1,58) = 6.776$, $p < 0.02$. The important comparison here is between subjects, so the significant differences between items are not relevant. These results therefore show that the LAA and MAA groups were comparable with respect to comprehension and recall of complex NP sentences.

As in Experiment II LAA performance, there were significant differences in MAA scores for different sentence segments in both non-verbatim memory, $F1(4,76) = 30.62$, $p < 0.001$; $F2(4,116) = 81.41$, $p < 0.001$ and verbatim memory, $F1(4,76) = 29.88$, $p < 0.001$; $F2(4,116) = 58.12$, $p < 0.001$ (Figures 9.13, 9.14). Comparing the MAA and HAA groups, the W-shaped recall curve is steeper for the MAA group, and there is a significant interaction of Group and Segment in both non-verbatim memory, $F1(4,76) = 4.51$, $p < 0.01$; $F2(4,116) = 15.11$, $p < 0.001$ and verbatim memory, $F1(4,76) = 4.96$, $p < 0.001$; $F2(4,116) = 13.16$, $p < 0.001$.

There were also group differences in the precision of recall. MAA levels of non-verbatim memory were lower than those of the HAA group. The interaction of scoring system and group was not significant by subjects, $F1(1,19) = 3.91$, $p < 0.0626$ but significant by items, $F2(1,29) = 19.50$, $p < 0.001$.

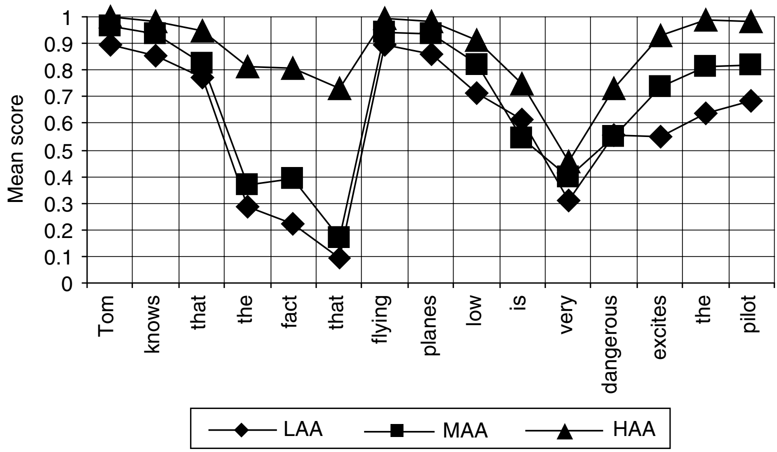

Figure 9.13 LAA, MAA and HAA recall curves by word

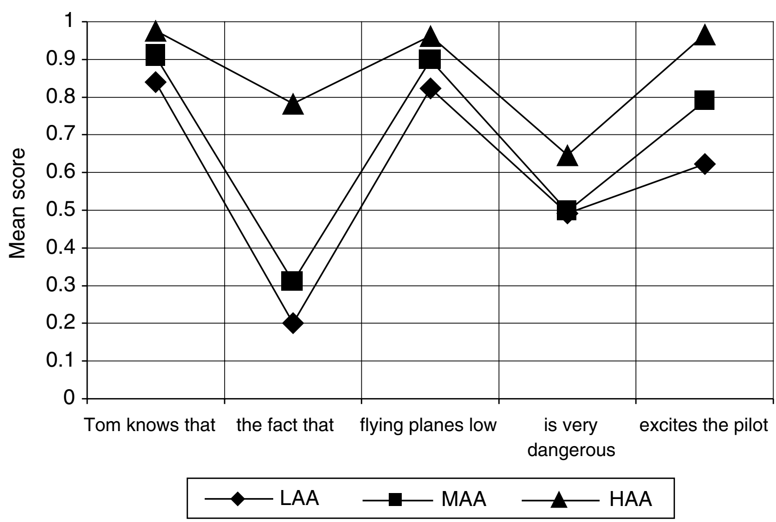

Figure 9.14 LAA, MAA and HAA recall curves by segment

Discussion

The results of the MAA group mirror those of the LAA group in Experiment I. When these two groups are compared, there was no significant difference in comprehension or recall scores. There were significant differences between comprehension and recall scores for the MAA and HAA groups. The recall curve for the MAA group was the same as that of the LAA group in Experiment I. Comprehension responses also resembled those of the LAA group, though these were not analysed in detail.

Phase 2

There was an effect of Phase on MAA comprehension scores, $F1(1,9) = 49.58$, $p < 0.001$ and by items, $F2(1,29) = 1626.1$, $p < 0.001$ (Figure 9.15). MAA comprehension scores exceed HAA comprehension scores after comprehension training. The interaction between Groups and Phase is significant, $F1(1,26) = 80.753$, $p < 0.001$; $F2(1,232) = 317.2$, $p < 0.001$ (Figure 9.15).

There was an effect of Phase for both non-verbatim memory, $F1(1,9) = 31.61$, $p < 0.001$; $F2(1,29) = 30.7971$, $p < 0.001$ and verbatim memory, $F1(1,9)=21.074$, $p<0.001$; $F2(1,29)=15.15$, $p<0.001$ (Figure 9.16). Comparing the HAA and MAA groups, there was an interaction of Group and Phase in non-verbatim memory, $F1(1,19) = 34.96$, $p < 0.001$; $F2(1,58) = 30.8$, $p < 0.001$, and verbatim memory $F1(1,19) = 15.17$, $p < 0.001$; $F2(1,58) = 15.15$, $p < 0.001$. Comparing the MAA and LAA groups in performance at pre-test (Experiment I for LAA) and post-test (Experiment II for LAA), there was a main effect of phase for non-verbatim

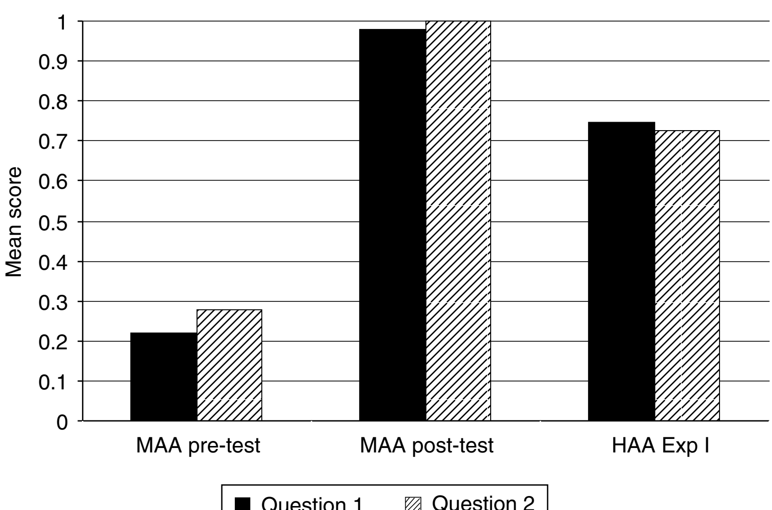

Figure 9.15 MAA and HAA comprehension scores after training

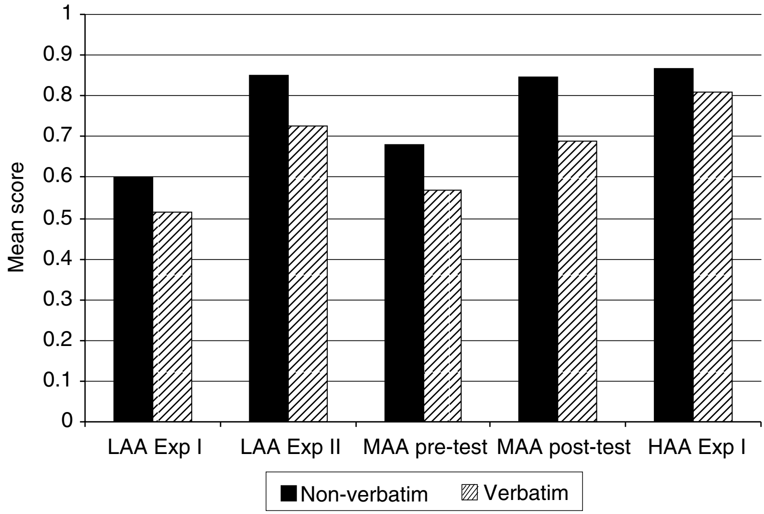

Figure 9.16 LAA, MAA and HAA recall scores

memory, $F1(1,26) = 81.094$, $p < 0.001$; $F2(1,58) = 137.438$, $p < 0.001$, and for verbatim memory, $F1(1,26) = 36.094$, $p < 0.001$; $F2(1,58) = 85.778$, $p < 0.001$, but there was no interaction between Group and Phase, for non-verbatim memory by subjects, $F1(1,26) = 1.576$, $p < 0.22$ (though significant by items, $F2(1,58) = 5.816$, $p < 0.02$) or for verbatim memory by subjects, $F1(1,26) = 0.79$, $p < 0.383$ (though again significant by items, $F2(1,58) = 6.544$, $p < 0.02$). In other words, memory and comprehension training had similar effects on recall scores.

The dips in the recall curve are now shallower than at pre-test (Figure 9.17) and the curve now closely resembles that of the HAA group. There was still an effect of segment for both non-verbatim memory, $F1(4,36) = 3.41$, $p < 0.02$; $F2(4,116) = 19.43$, $p < 0.001$ and verbatim memory, $F1(4,36) = 4.28$, $p < 0.01$; $F2(4,116) = 13.7930$, $p < 0.001$. However, there was no significant interaction between Groups (MAA and HAA) and Segment for verbatim memory, $F1(4,76) = 1.4$, $p < 0.24$, $F2(4,232) = 2.01$, $p < 0.09$ or for non-verbatim memory by subjects, $F1(4,76) = 2$, $p < 0.1$ but the interaction was significant by items, $F2(4,232) = 2.95$, $p < 0.02$.

Discussion

Comprehension training led to an improvement in recall comparable in magnitude to the improvement caused by recall training. This finding is

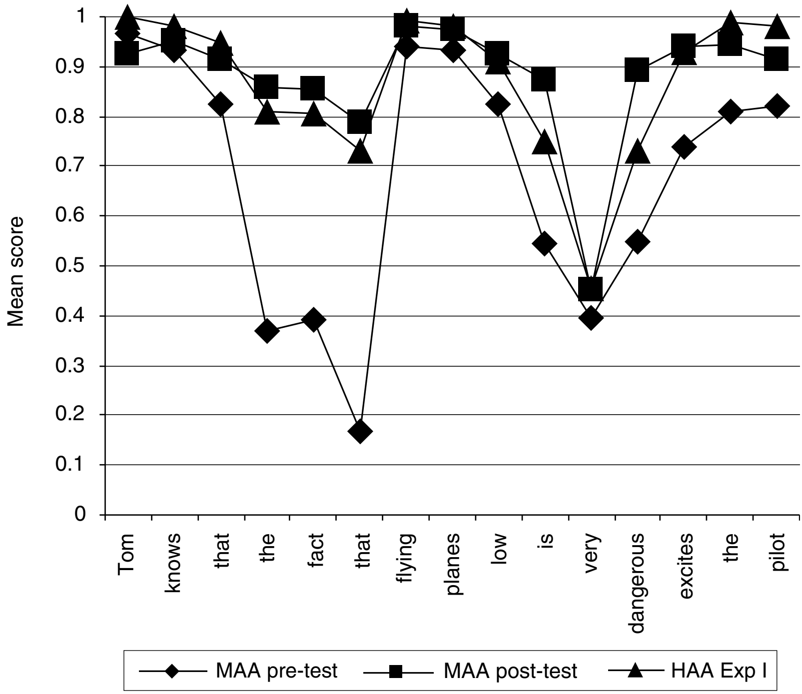

Figure 9.17 MAA and HAA recall curves by word

consistent with Ericsson and Kintsch's theory that sentence encoding operations have a bearing on recall. It is not clear what Just and Carpenter's theory has to say about effects of comprehension training. The results of this experiment therefore do not have a direct bearing on their theory.

General discussion

The experiments described in this chapter investigated the comprehension of complex NP sentences of the form: *Tom knows that the fact that flying planes low is dangerous excites the pilot*. Steps were taken to ensure the grammaticality of the test sentences. They were constructed through the recursive elaboration of a simpler form, such as, *Tom knows that flying planes low is dangerous*. The complex form involves no rules which are not present in the simpler form, except for the fact of recursion itself. The complex form was also approved by two expert syntacticians. Further, there are many examples of this sentence form in the British National Corpus, many of which are far more complex than the sentences used in the experiments. Failure to comprehend sentences of this form cannot therefore be attributed to ungrammaticality. Alternatively, to say that the sentence is ungrammatical for some speakers of English but not others is to agree that native speakers differ in syntactic representation.

The experiments showed that native speakers of different educational backgrounds differ in their ability to understand these sentences, though a high level of education does not guarantee successful comprehension. There is also evidence to suggest that non-native speakers of English with expertise in linguistics can comprehend these sentences better than native speakers with a high school-only level of education. It remains to be seen whether highly educated non-native speakers of English with no expertise in linguistics can also comprehend such sentences better than native speakers educated only up to high school level. It was suggested that the higher performance of the non-native speakers may have been due to both linguistics expertise and the fact of having learnt English syntax through explicit instruction.

The experiments also showed that native speakers of the same age but with differing levels of academic achievement also understand these sentences differently. The experiments did not seek to uncover the ultimate source of these differences. Rather, the experiment sought to decide whether these differences can be explained in terms of variations in working-memory capacity or in terms of variations in syntactic representation. The experiments suggested a certain relationship between comprehension and recall. The performance of the high school-only native speakers in the preliminary experiment suggested that they were imposing a syntactic formula on the sentences in a top-down fashion and ignoring vital syntactic details. The recall performance of students with low and medium levels of academic achievement revealed a pattern of structural forgetting consistent with the imposition of such a schema. That is to say, the parts of the sentence which these subjects recalled were just the ones that could be fitted into the schema whereas the parts that they forgot were just the ones that could not be fitted into the schema. Further work is needed to determine if this interpretation is correct, as no attempt was made to correlate comprehension responses with recall curves in Experiment I. It is also possible that the individuals who comprehended the test sentences correctly might simply have imposed a more complex syntactic formula. However, given the rarity of the construction used in the experiment (see Granath, 2001), this possibility is not very strong.

Training subjects to recall the test sentences resulted in improved recall but there were no benefits for comprehension. Training subjects to understand the test sentences, however, led to improvements in both comprehension and recall. This pattern of results is not consistent with the idea that the individual differences observed in these experiments were due to individual variations in working-memory capacity. If that had been the case, then one would have expected recall training to boost the amount of activation available for comprehension, leading to improvements in comprehension. In fact, there was a slight and inexplicable decrease in comprehension performance after recall training. On the other hand, the improvement in both comprehension and recall due to comprehension is consistent with the idea that comprehension depends on the ability to represent sentences in an efficient, task-relevant manner in LTM. This efficient form of representation would facilitate both comprehension and recall. In that case, the individual differences observed in Experiment I could be explained in terms of individual differences in syntactic representation rather than inherent individual differences in working-memory capacity.

There were some other interesting patterns of individual differences which require further investigation. The paraphrase task during comprehension training in Experiment III showed that LAA subjects could understand the test sentences when they were presented in a different form. This suggests that the complex NP is harder to process only if it occurs in an unfamiliar context. There is evidence from research in language development that children find complex NPs easier to process if they occur pre- rather than post-verbally. It is difficult to say if this was the case with the present subjects because the experiment did not set out to vary the syntactic context of the NP. If it is the case that context matters, then it would suggest that some individuals have overly context-dependent representations of nouns, so that only nouns in certain contexts can be post-modified by a clause. Hopefully, further research will clarify the issue.

Secondly, LAA subjects displayed lower levels of precision in recall. This might suggest that their lexical representations are less precise than those of HAA subjects. Again, it is impossible to be sure because the experiment was not set up to examine these issues. Nevertheless, it is noteworthy that Cupples and Holmes (1992) also found individual differences in the precision of lexical knowledge. In their study, less skilled readers performed poorly in assigning words to lexical categories. Cupples and Holmes suggested that poor lexical representations lead to a reduced ability to assign phrasal organisation to sentences. They suggested that individuals with poor lexical representations may not perform syntactic analysis at all, but simply use content words to decide on the most probable interpretation of a sentence. Cupples and Holmes noted that these subjects were also adversely affected by sentences involving semantic reversibility between the noun concepts.

In line with this possibility, several subjects in the current experiments commented that the sentences were ambiguous or that they had several possible interpretations. Given that the sentences are not ambiguous, it is possible that subjects thought they were, simply because they were not using syntactic cues to work out the grammatical relations between sentence constituents. Rather, as suggested by Cupples and Holmes, they may been using the content words to work out the most likely interpretation. The pattern of structural forgetting, however, suggests that subjects may have resorted to a semantic strategy only after failing to comprehend the test sentences by using a syntactic formula.

A third pattern of individual differences took the form of differential rates of transfer of training to test. Although all subjects had learnt to paraphrase the sentences correctly during training, they differed in their ability to transfer what they had learned to the different task of answering comprehension questions. In the previous chapter, it was suggested that individuals may differ in inductive ability, that is, to see the similarities between diverse phenomena. In this instance, many subjects failed to notice that the sentence representation generated during the paraphrase task could also be used to answer comprehension questions. Subjects had to be told to transfer training to test in order to perform well. A potentially fruitful line of investigation might be to study individual differences in inductive ability.

9.4 Conclusion

The overall pattern of results from all four experiments is generally in favour in Ericsson and Kintsch's theory. The pattern of structural forgetting observed in Experiments I and IV cannot be explained by Just and Carpenter's theory, but it is consistent with the idea that sentence encoding operations impact on recall. Experiment II found that recall training improved recall but did not affect comprehension. This is unexpected from the perspective of capacity constrained comprehension, which suggests that freeing capacity from storage tasks should lead to improvements in processing. Experiment III showed that comprehension training leads to improvements in comprehension. This finding is consistent with the presence of a syntactic deficit and is hard to explain in terms of a working-memory deficit. Experiment IV showed that comprehension training leads to improved recall, a finding which is consistent with the idea that efficient sentence encoding facilitates recall performance.

The individual differences observed in the experiments can be described in terms of individual differences in the definition of syntactic units. Some individuals appeared to impose an inappropriate syntactic formula on the input in a top-down fashion. Other individuals appeared to carry out detailed syntactic analysis in a bottom-up fashion. These individual differences correspond to expert–novice differences described in Chapter 1. They also correspond somewhat to experience-based versus rule-based forms of syntactic processing. The experimental results reported here therefore provide further support for the need to integrate experience-based and rule-based approaches to language.

10
Implications

10.1 Introduction

The broad aim of the research described in this book was to decide between rule-based and experience-based approaches to the description of linguistic knowledge. This goal was approached in terms of a more specific research question, which was to decide between two alternative accounts of variations among native English speakers in the ability to understand complex sentences. During the course of the research, it became apparent that neither approach is sufficient by itself. Language users display capacities for both associative learning and rule-following. Language users also appear to differ in the extent to which they manifest each of these capacities. It therefore became not a question of deciding between the two approaches but of finding appropriate ways to integrate them. An integration was proposed based on psychological and linguistic considerations. The integrated approach is for the moment too broad to constitute a model of language processing. However, it is consistent with a number of experimental findings and it should yield in due course to mathematical formalisation. This chapter will briefly recapitulate the major arguments and suggest future developments. It will then explore some pedagogical implications and potential applications of the research.

This chapter is divided into two sections. The first section is largely theoretical while the second is concerned with pedagogical issues.

10.2 Theoretical issues

This book began by describing two basic approaches to the description of language: the experience-based and the rule-based approaches. The experience-based approach, in its early and current forms, describes language in terms of sequential probabilities. This approach is supported by two major observations: that natural language displays statistical regularities and that language users are sensitive to these regularities. However, the experience-based approach underestimates what language users know about their language. Contiguity and frequency alone are insufficient to describe knowledge of language. There is evidence from the psychology of skill acquisition to suggest that human beings are capable of constructing representations whose complexity far exceeds the limits of sequence learning. The rule-based approach is therefore correct to attribute language users with the potential for rule-guided behaviour. However, the rule-based approach overestimates what native speakers know about their language. Language users do not necessarily use linguistic rules. In fact, language users appear to differ in the extent to which they rely on statistical or rule-guided processes in language comprehension.

A more comprehensive account of individual differences was proposed which incorporates psychological and linguistic perspectives. Ericsson and Kintsch's (1995) theory of skilled processing combines the ability to learn from experience with the ability to follow rules. According to that theory, individuals engage in control processes in order to construct task-specific systems of representation. These systems of representation both capture the abstract structure of a given task as well as anticipate its informational requirements. An important aspect of the theory is that the systems of representation are constructed in long-term memory (LTM) and they effectively extend total working-memory capacity. The degree of an expertise in a given domain correlates positively with the efficiency with which these representational systems utilise LTM. It is therefore not the case that the capacity of working memory constrains cognitive processes. Rather the reverse is true: it is the efficiency of cognitive representations which determines the total capacity of working memory. The theory therefore undermines the competence–performance distinction by locating cognitive constraints in the quality of representations as opposed to the computational resources required to construct those representations.

Saussure's theory also integrates rule-based and experience-based perspectives from a linguistic angle. The theory states that human beings have an innate capacity to *construct* languages. Linguistic systems emerge inductively from experience. They also incorporate their own productions: expressions that are generated by the system are stored in LTM and become part of the system, just as mathematical formulae comprise an intrinsic part of mathematical knowledge. According to the theory, the construction of language is based on four interacting laws of association. It was argued that these laws subsume both associative learning and rule-following. The law of contrast creates distinct units through systems of contrasts. The law of similarity creates classes of units and enables rule-guided behaviour. The law of contiguity governs the sequential arrangement of units. The law of frequency sharpens or reduces contrasts, influences the induction of categories and strengthens or weakens sequential associations. The interaction of these laws defines the basic linguistic units employed by an individual.

Evidence was presented to suggest that a wide range of individual differences can be described in terms of individual differences in the definition of linguistic units. The optimum unit depends on the characteristics of the linguistic input and the task. Units which occur frequently in a sequence may become unitised. This unitisation increases cognitive efficiency by enabling several items to be treated as a single item. However, unitisation obscures details of internal structure. Consequently, system-wide patterns of similarity and contrast are also obscured. On the other hand, if the constituents of a sequence occur frequently in a wide range of other sequences, then differentiation may occur as the contrasts between constituents are sharpened. Details of internal structure emerge, revealing system-wide patterns of similarity and contrast. To say that individuals differ in the definition of their linguistic units is therefore equivalent to saying that they differ in their linguistic systems. Saussure emphasised this interdependence between unit and system strongly.

In certain tasks, individuals who employ unitised word sequences in a holistic manner appear not to engage in detailed bottom-up analysis of the linguistic input. Instead, they operate in a top-down fashion, even if this sometimes leads to error. This behaviour is consistent with an experience-based approach in the sense that it involves finding the best fit to the input from a stock of learned syntactic patterns. On the other hand, individuals with more analytical units appear to engage in detailed bottom-up analysis, leading to the construction of accurate representations. This behaviour is consistent with a rule-based approach in the sense that complex representations are derived deductively as opposed to being selected from a stock of prefabricated representations.

It is not the case, however, that small is good and big is bad. In text processing, for instance, less skilled readers construct small and isolated text units whereas skilled readers construct more global text units. As stated in the previous paragraph, the optimum unit depends on the nature of the input and the task. It should also be emphasised that it is not merely a question of unit size but of the precise manner in which each unit is defined, that is, the system-wide pattern of contrasts and similarities.

In the case of syntactic processing, it appears that some individuals impose syntactic formulae on the input. These formulae may be appropriate for frequently occurring patterns of linguistic input. However, confusion can result when the formulae provide a poor fit for unusual inputs. Some evidence was presented to suggest that the use of formulae can be detected in patterns of comprehension and recall. Another group of individuals appears to generate global syntactic representations by combining appropriately defined units. These individuals behave in a more flexible manner and are able to understand complex sentences correctly. The pattern of individual differences just described was observed in an experiment carried out to decide between accounts of individual differences based either on variations in working-memory capacity or variations in syntactic representation. The experiment supported the latter account. Individuals who had initially failed to comprehend or recall test sentences were able to do so after receiving appropriate training. Given that the capacity of working memory is thought to be fixed, the effect of training must have been to alter the kinds of representations used by the poor comprehenders.

Saussure and Ericsson and Kintsch's theories have positive implications for language pedagogy and for human mental potential in general. Ericsson and Kintsch's theory locates the source of cognitive constraint in systems of representation rather than in computational resources. Furthermore, the theory proposes that human computational capacity can be amplified by increasing the efficacy of the representational systems. Saussure's theory, on the other hand, proposes that human beings are innately equipped with the capacity to create systems of representation. The theory also provides a general description of the organisation of such systems. The combination of the two theories suggests that human cognitive limits are defined by the systems of representation, or languages, which underlie cognition, rather than by inherent computational limits. There is tangible support for this conclusion. The development of mathematics has amplified human thought to an immeasurable degree.

It is interesting to note, in view of the last comment, that the theories of Saussure and Ericsson and Kintsh are amenable to a treatment in terms of a branch of mathematics called graph theory. In fact, graph theory has a long history in linguistics and psychology. I provide a brief review of the applications of graph theory in these two disciplines below.

As I have already related in earlier chapters, the idea that language is a network was first proposed by Ferdinand de Saussure (1916). This view was taken up subsequently by Hockett (1955) and Lamb (1966). Currently, the notion of language as a network is advocated by Goldberg (1995), Halliday (1985), Hudson (2000) and Langacre (1990). In spite of this long history, however, linguists have not been able to take the network view of language very far, apart from approaches such as Lamb's Stratificational Grammar and Hudson's Word Grammar. The failure to capitalise on the insight that language is a network arises largely because linguistics does not, in itself, have the general conceptual tools needed to describe networks. Such tools are found in graph theory but they are largely inaccessible to many linguists, who lack the requisite grounding in mathematics.

A number of mathematicians with an interest in graph theory, have, however, applied graph theory to linguistic analysis. For instance, Markov (1913) developed the seminal theory of Markov models from a study of Pushkin's novel – *Eugene Onegin*. Polya (1954) and Hubey (1999) provide highly informative applications of graph theory to the study of language families. Textual analysis has been another fruitful domain of application for graph theory (for example, Auster, 1980; Boot, 1976; Dailey, 1959; Quentin, 1926; Zarri, 1976).

Graph theory has also been applied to natural language in the form of Markov models. A discrete Markov model specifies a number of states in terms of a *state vector* and a set of transition probabilities in terms of a *transition matrix*. Multiplying the state vector at one point in time by the transition matrix gives the state of a system at the next point in time. Iterating the process gives a sequence of states which describe the behaviour of a system in time. Depending on the starting state and/or the transition matrix, this behaviour can reach a steady state from which the system cannot escape. Shannon and Weaver (1949) used Markov models to describe the statistical structure of English text. Subsequently, Markov models have been applied to the description of the statistical structure of natural languages for the purposes of creating computer programs that can process natural language in one way or another (for example, Forney, 1973; Garside, 1987; Rabiner, 1989; Sharman, 1989).

Another domain of application for graph theory has been in psychology. Psychologists have long held that knowledge is represented in memory in the form of associative networks and, to the extent that language is a form of knowledge, it too has the structure of a network. Kiss (1968) used graph theory to study lexical networks. He also suggested the use of Markov chains and signal-flow graphs to describe stimulus–response chains. Frase (1969) used concepts from graph theory to study the representation of texts in memory while Greeno (1976) used relational network theory and graph theory to describe the mental representation of structured knowledge. Costa-Pereira and Maskill (1983) analysed essays by students writing on a newly learned topic in chemistry using concepts from graph theory. Ross (1994) and Sutton *et al.* (1994) use graph theory to describe the sequencing of elements in poems and narratives produced by normal individuals and victims of mental disorder. Wilks (1999) has applied graph theory successfully to the study of the mental lexicon.

Graph theory is used widely in psychology in the form of connectionism. Connectionist models can be thought of as Markov models adapted in order to simulate psychological processes (see McLelland and Rumelhart, 1986; Rumelhart and McLelland, 1986). As related in earlier chapters, connectionist models have been used to simulate language learning and sentence comprehension (for example, Christiansen and Chater, 1999; Elman, 1993).

This brief review demonstrates the relevance of graph theory to both linguistics and psychology. It appears that graph theory can seal two sorts of rifts. On the one hand, as argued earlier, it makes it possible to combine rule-based and experience-based approaches to language and linguistic behaviour. On the other hand, it also makes it possible to harmonise the relationship between linguistics and psychology by providing a technical language which can be used to describe both linguistic representations in the abstract as well as actual psycholinguistic structures and processes. The use of graph theory can also bring linguistic and psychological research into alignment with more established sciences, as argued below.

Work by the physicist Barabási (2002) indicates that networks are ubiquitous in nature. It appears that networks have similar structures whether they are found in molecules, social relationships or even the world-wide web. Studying language as a network therefore has the potential to reproduce results that have been found in other domains. Such a development would be beneficial for language studies, which are currently isolated from other fields of research. The sharing of results and techniques with more established sciences would also help establish the scientific approach in linguistics.

The foregoing demonstrates the relevance of graph theory to the study of language. However, the lack of mathematical expertise has prevented many language researchers from making use of it. In addition, the application of graph theory has been rather piece-meal and many researchers are only familiar with specific applications. Many are unaware of the overarching theory and the broad range of descriptive concepts which it provides. Language researchers are therefore faced with the challenge of learning graph theory and associated mathematical systems in order to make substantial advances in the field. This will be an uphill struggle because many language researchers do not have a background in mathematics. I have personally begun to correct this serious shortcoming in my education and I am currently taking courses in mathematics.

Having explored some theoretical consequences of the ideas presented in this book, I will now turn to some pedagogical applications. Although there are some potential applications of graph theory to language teaching, I will confine myself to Saussure's four principles of similiarity, contiguity, contrast and frequency. Elements of graph theory are, of course, implicit in these principles.

10.3 Pedagogical implications and applications

A major pedagogical implication of the discussion is that it is necessary to provide instruction in the grammar of the first language during the school years. Indications are that being a native speaker of a language does not automatically mean that individuals will be able to construct linguistic systems which enable them to understand, at a syntactic level, all the sentences of that language. A second implication is that this instruction, if appropriate, will enhance linguistic performance. While both experience-based and rule-based approaches impose strict limits on human linguistic abilities, the points of view developed by Ericsson and Kintsch, and Saussure suggest that limits in linguistic performance are imposed only by the nature of the representations employed by an individual. Given that these representations are susceptible to modification, it stands to reason that appropriate instruction will result in improved performance.

Language pedagogy has been subject to a dichotomy similar to that between rule-based and experience-based approaches. There is a tension between the teaching of rules and the teaching of language in use, and the history of language teaching is marked by swings between one extreme or the other. The foregoing discussion indicates that language pedagogy should incorporate both rule-based and experience-based approaches. The foregoing discussion also indicates the importance of taking into consideration the psychological findings on skill acquisition. The following paragraphs will discuss some potential pedagogical applications of this integrated approach in terms of the laws of contrast, similarity, contiguity and frequency. Some applications are discussed in relation to the various experimental techniques and results described in several parts of the book. An important point to bear in mind in considering the following suggestions is that they are meant to be scalable to every sort of linguistic unit. Thus suggestions about spelling can form the basis for exploring similar applications to syntax or the text level. Many of the suggestions are also made with the aim of fostering awareness of linguistic regularities in an implicit fashion which is suitable for young children. Some of the suggestions are also applicable to adult learners.

The following discussion does not refer to the literature of language teaching and learning. This would be too great an undertaking and it would be to stray too far from the focus of this book. Some of the applications described below are not necessarily novel and may already be in practice. The aim of the discussion is simply to *illustrate* the ways in which the four associative principles can be put to use for pedagogical purposes. The potential applications are therefore presented suggestively. Ingenuity and experimentation would be required to make them practicable and effective.

10.4 Contrast

Contrast is the linguistic principle which defines individual units by distinguishing them from each other. It is also the psychological principle which is necessary to counteract the effects of interference in recalling information from LTM. One way of applying this principle is to make significant contrasts salient in the input in order to facilitate recall of linguistic forms during production. It was mentioned earlier that, when sequences are unitised, the contrasts between their constituents may be lost. For instance, students may learn to recognise whole words on sight during reading, but they may not be able to spell those words correctly themselves because they have not paid sufficient attention to the orthographic structure of each word.

This observation raises the question of whether their spelling might improve if the individual letters in each word were made more distinct from each other by manipulating fonts. For instance, if the word ‘favourite’, which young children find difficult to spell, were printed as ‘favourite’, this might encourage them to attend to the orthographic structure of the word and thereby make it memorable. If computer displays are used, it might be possible to animate words by using different fonts for adjacent letters and alternating these fonts periodically between the two letters. Another possibility is to get children to write words using different colours for each letter. This would add the dimension of colour to the contrasts between individual letters. Another possibility would be to manipulate fonts in order to emphasise the differences between homophones, for instance, '**t***h***e***r***e**' versus '*t***he***i***r**'. The idea is that frequent exposure to print which emphasises relevant contrasts might lead to the formation of distinct representations for otherwise confusable items. Care would need to be taken, of course, to ensure that children do not start to use different fonts within each word.

This idea can be scaled to other kinds of linguistic units. A suggestion will be made later in the chapter to use different fonts or colours for different lexical classes. In addition, contrasts based on colour or font could be used to mark different text segments in order to foster an awareness of text structure. For instance, various parts of a narrative, such as beginning, middle and end could be printed in different colours. Different textual functions, such as describing, explaining, exemplifying and so on could also be marked in the same way.

There are many possible variations of applying the basic principle of contrast. One possibility might be to design a computer game which involves episodes in which the player must navigate through a complicated building via a series of locked doors. Each door can be unlocked by keying in a code. The code for each door would be a letter which comprises part of an often misspelled word, such as 'favourite'. The player could be told verbally what the whole word is beforehand, but he or she would have to key in each letter by themselves in the correct sequence, in order to test or foster their spelling ability. Dire consequences would befall a player who entered the wrong code. This idea could also be applied in order to foster sensitivity to orthographic probabilities by using sequences of codes which mimic the probability of one letter following another. There could be desirable outcomes if a player's codes follow the most probable sequence and undesirable outcomes if the player's codes follow less probable sequences.

The codes in the game suggested above need not be letters but could be any kind of linguistic unit. They could be sequences of words which comprise a grammatical sentence. They could be sequences of events. Stories tend to follow certain patterns according to a sort of narrative grammar. Therefore, instead of being asked to key in a letter, the player might be presented with various predictions about what is going to happen next, with some predictions being more probable than others. Fostering the ability to predict future events in developing narrative would enhance the learner's knowledge of narrative grammars. This, in turn, could help them in reading or in writing their own narratives.

10.5 Similarity

The principle of similarity is related to the formation of categories. The categories, in turn are related to the formation of sequences. Evidence was presented in Chapter 8 which suggested that some native speakers of English cannot assign words to parts of speech accurately, and that their ability to construct syntactic representations suffers as a result. It might be possible to design reading materials in which words from different word classes are printed in distinctive fonts or colours. Over time, students could begin to recognise sets of words based on font or colour. In addition, this form of presentation would give students a visual impression of the distribution of word classes in a clause. Theoretically, it would then be easier for students to appreciate the notion of clause if they recognise that clauses contain words of certain colours in certain sequences. The idea is to maximise the number of cues to linguistic structure. It would be important, of course, to ensure that learners do not become too dependent on visual cues. The idea is to lead them to an awareness of the abstract *linguistic* structure, so they would have to be weaned from non-essential cues at some point.

This idea could be generalised to the production of materials in other areas of the curriculum. For instance, in mathematics, the notions of percentage, proportion, ratio, gradient and derivative can be considered as instances of a more general concept, in the sense that understanding one of these concepts helps to understand the others. If the words percentage, proportion, ratio, gradient and derivative were to be colour-coded, this might facilitate student's reading and understanding of mathematical texts. This is because the colour coding might help activate other concepts with the same code. The conceptual structures associated with these activated concepts might then be brought to bear in the understanding of a new but related concept. Of course, texts will often make explicit reference to related concepts in explaining a new concept. However, there may be circumstances where it might be useful to have implicit references.

Chapter 3 also briefly reviewed some studies which indicate that layout can be used to help readers impose phrasal organisation. For instance, Anglin and Miller (1968) found better recall for passages that were segmented in conformity with phrase structure compared to passages whose segmentation violated phrase structure. Graf and Torrey (1966) also found better comprehension when rapidly presented text segments conformed to major rather than minor syntactic boundaries. Text layout could therefore be exploited to facilitate the perception of phrase boundaries. It should also be possible to teach phrase structure explicitly, once the notion of lexical class is well established. Phrase structure is not complicated and it has a regular structure which children should be able to appreciate.

10.6 Contiguity

Contiguity relates to the linear arrangement of units. An important aspect of contiguous relations is the way in which the context in which an item occurs influences its role in a sequence. Context effects can be observed in the following contrived sentence: 'Buffalo buffalo buffalo buffalo buffalo'. This sentence shows how a syntactic formula can be imposed on a string of words in a manner which gives each word a unique role. In the example, the first use of buffalo is to modify the head noun, which is also buffalo. The third use of the word is as a verb and the last two uses are as modifier and as head noun. Thus the same lexical item can be assigned to several roles depending on its position in a sentence. Chapter 7 discussed the idea that sentence comprehension can also be a top-down process whereby a syntactic formula is imposed on the input. Although the use of such formulae can sometimes have negative effects, as described earlier, it should be recognised that their use can also be creative, in that it can create new senses of words. An example of this is provided by the Geer *et al.* (1971) study, which showed that graduate native English speakers could understand that 'boot' in 'boot green' is a modifier, whereas the high school-only native English speakers insisted on interpreting the phrase as 'green boot'. It might therefore be useful to make students aware of the fact that new senses of words can be created by placing them in new contexts.

Contiguity also relates to sequences beyond the sentence. Work was reviewed in Chapter 7 which showed that less-skilled readers construct smaller textual units than skilled readers. One manifestation of the inability to construct integrated text representations is the difficulty experienced by some learners to resolve anaphoric references if the distance between anaphor and antecedent is too great (see Oakhill, 1994). One way of addressing this problem would be to use text layout in order to make textual relations more salient. For instance, chains of reference could be marked by specific fonts, in order to allow each discourse referent to be tracked visually. If text is presented via computer screen, then elements in each chain of reference could be associated via hyperlinks. Learners could be asked comprehension questions which require them to follow a chain of reference via the hyperlinks, the aim being to get them to appreciate large-scale patterns of text and the need to create more integrated text representations.

10.7 Frequency

Frequency is closely related to the other three laws. A major potential application of this law is to automate various language tasks. Chapter 6 described some experiments by Mowbray and Rhoades (1959) which showed that reaction time decreases with practice. More recent experiments by Logan (1988) also show that practice leads to significant decreases in both the time taken to perform simple tasks and the number of errors produced. Crutcher (1990, 1992) found that practice reduces the retrieval time for paired associated by a half. In general, the finding is that practice increases speed and reduces error by a power function. There are large gains at the beginning of training which decrease gradually to asymptotic levels. The power law of learning is apparently one of the most well-established laws in psychology and it makes sense to make more systematic use of it for those aspects of linguistic skill which are susceptible to automatisation.

One application would be to enhance speed of item recognition. For instance, very young children who are still learning the alphabet could benefit from being able to recognise letters rapidly. The recognition of other stable units, such as words could also be enhanced by tasks involving rapid reaction to stimuli. One task could involve the ability to discriminate rapidly between correct and incorrect spellings of a word. For instance, correct or incorrect spellings could be flashed on a screen and learners would have to press a given key if the spelling is correct and a different key if it is incorrect. The computer could measure reaction times and repeat the presentation of certain items until recognition speed reaches desired levels. The computer could also provide instant feedback in the form of tones or recorded messages. The task would be masked in some game scenario where winning requires rapid reaction.

A variation of this task could involve lexical class assignment. Learners could be presented with a sentence on the screen in which one of the words is missing. The computer could then present a series of possible candidates and the task would be for the learner to press a given key once a suitable candidate appears. In this case, reaction time would be a measure of the degree of automaticity in the lexical assignment component of syntactic processing. Reaction time techniques could also be used in other areas of the curriculum as an effective method of memorising key facts.

Chapter 3 reviewed studies which indicate that language users are sensitive to the statistical structure of language. These studies involved Hockett's (1953) guessing game, which shows that letter guessing is more correct within words than between words. A number of other studies showed that predictability influences performance in tasks such as recall (Deese and Kaufman, 1957; Marks and Jack, 1952; Miller and Selfridge, 1951; Richardson and Voss, 1960; Sharp, 1958), language perception (Miller *et al.*, 1954; Pollack and Pickett, 1964; Traul and Black, 1965), language production (Goldman-Eisler, 1958; Maclay and Osgood, 1959) and attention span (Imae and Takeuchi, 1959). Some studies indicated that skilled readers are more sensitive to transitional probabilities in orthography than less skilled individuals (for example, Lefton *et al.*, 1973; Scheerer-Neumann *et al.*, 1978). Experiments by Perlmutter and MacDonald (1995) also show that readers with different levels of skills differ in their sensitivity to lexical probabilities. Learners should therefore benefit from tasks which increase their sensitivity to sequential probabilities.

This sensitivity can only be built up over long periods of time. It could therefore be enhanced by designing reading materials which make certain sequential patterns more salient. These patterns could be determined beforehand using corpus analysis techniques. It might also be useful to design guessing games (for instance, Hockett, 1953) which encourage learners to pay more attention to sequential pattern in their everyday use of language. Other games could involve attempting to decipher messages that are degraded in one way or another. For instance, some letters or words may be missing or indistinct, and learners would have to guess the missing material from the context, rather like the cloze task. The important point would be to ensure that the materials represent certain statistical regularities in language.

10.8 Manipulating print

A number of the proposals suggested above involve the manipulation of print. The basic idea is that print can be brought to the service of pedagogy in a more systematic and perhaps more effective manner.

This idea treats the various dimensions of print, such as font, layout and colour as a visual language for representing more abstract conceptual relationships. Print, after all, is a form of representation. Maximum benefit can be derived from it only if it is treated as a language in its own right, having the same basic structure as other languages. In particular, it has to be systematic and it has to avoid interference. Considerable thought and experimentation would therefore need to go into designing effective schemes for exploiting print.

The various proposals made above are meant to illustrate possible ways of applying linguistic and psychological principles to language learning. The proposals do not exhaust the possibilities, but it is hoped they are sufficiently suggestive. A far more general pedagogical implication of the research is suggested below.

10.9 Conclusion

This book has indicated the centrality of language for cognition. The word 'language' is used loosely here to refer to systematic forms of representation created in the service of cognitive tasks. It has been proposed, on the basis of Ericsson and Kintsch's theory, that cognition depends on the development of domain-specific systems of representation. Saussure proposes that the representational systems constructed by the human mind have a common form of organisation and can be subsumed under a more general science, which he called semiology, or the science of signs. If both proposals are correct, then they lead to the thought that different areas of the curriculum should be regarded as systems of representation with the same basic organisation as that of language. If so, then learning any subject involves learning a language. This formulation has the advantage of generalisation. It suggests that all learning involves essentially the same kind of underlying activity. What we know about linguistic structures and processes can therefore be transferred to other domains of learning. The potential outcome is a much more systematic approach to education.

References

Aborn, M. and Rubenstein, H. 1952. 'Information Theory and Immediate Recall', *Journal of Experimental Psychology, 44*, 260–266.
Anderson, J. R. 1974a. 'Verbatim and propositional representation of sentences in immediate and long-term memory', *Journal of Verbal Learning and Verbal Behaviour, Apr., 13/2*, 149–162.
—— 1974b. 'Retrieval of propositional information from long-term memory', *Cognitive Psychology, 6*, 451–474.
—— and Bower, G. H. 1980. *Human Associative Memory*. New Jersey: Lawrence Erlbaum.
Anglin, J. M. and Miller, G. A. 1968. 'The role of phrase structure in the recall of meaningful verbal material', *Psychonomic Science, 10/10*, 343–344.
Atkinson, R. C. and Schiffrin, R. M. 1968. 'Human memory: A proposed system and its control processes', in K. Spence and J. Spence (eds). *The Psychology of Learning and Motivation, 2*, 89–195. New York: Academic Press.
Auster, C. J. 1980. 'Balance theory and other extra-balance properties: An application to fairy tales', *Psychological Reports, 47/1*, 183–188.
Baddeley, A. D. 1986. *Working Memory*. New York: OUP.
—— 1990. *Human Memory: Theory and Practice*. Boston, MA: Allyn and Bacon.
—— and Hitch, G. J. 1974. 'Working memory', in G. H. Bower (ed.). *The Psychology of Learning and Motivation, 8*, 47–90. New York: Academic Press.
—— Thompson, N. and Buchanan, M. 1975. *Journal of Verbal Learning and Verbal Behaviour, 14/6*, 575–589.
Barabási, A. 2002. *Linked: The New Science of Networks*. Cambridge, MA: Perseus Publishing.
Baruzzi, A. 1982. 'Effects of Degree of Education on the Comprehension of Syntactic Structures in Normal and Aphasic Populations', *McGill Working Papers in Linguistics, 2*, 56–74.
Bates, E., Bretherton, I. and Snyder, L. 1988. *From First Words To Grammar*. Cambridge: CUP.
—— Masling, M. and Kintsch, W. 1978. 'Recognition memory for aspects of dialogue', *Journal of Experimental Psychology: Human Learning and Memory, 4/3*, 187–197.
Berko, J. 1958. 'The child's learning of English morphology', *Word, 14*, 150–177.
Bever, T. G. 1970. 'The cognitive basis for linguistic structures', in J. R. Hayes (ed.). *Cognition and the Development of Language*, 279–352. New York: Wiley.
—— 1988. 'The psychological reality of grammar: a student's eye-view of cognitive science', in W. Hirst (ed.). *The Making of Cognitive Science: Essays in Honour of George A. Miller*, 112–142. Cambridge: CUP.
—— Kirk, R. and Lackner, J. 1969. 'An autonomic reflection of syntactic structure', *Neuropsychologia, 7/1*, 23–28.
—— Lackner, J. R. and Kirk, R. 1969. 'The underlying structures of sentences are the primary units of immediate speech processing', *Perception and Psychophysics, 5/4*, 225–234.
Biber, D. 1983. 'Differential Competence in Somali: Evidence from the acquisition of Noun Definitisation', *Journal of Psycholinguistic Research, 12/3*, 275–295.

Blaubergs, M. and Braine, M. 1974. 'Short-term memory limitations on decoding self-embedded sentences', *Journal of Experimental Psychology, 102/4*, 745–774.

Bloomfield, L. 1926. 'A set of postulates for the science of language', *Language, 2*, 153–164. Cited in P. H. Matthews. 2001. *A Short History of Structural Linguistics*. Cambridge: CUP.

—— 1927. 'Literate and Illiterate Speech', *American Speech, 2*, 432–439.

—— 1933. *Language*. Chicago: University of Chicago Press.

—— 1936. 'Language or ideas?', *Language, 12*, 89–95. Cited in P. H. Matthews. 1993. *Grammatical Theory in the United States from Bloomfield to Chomsky*. Cambridge: CUP.

Blumenthal, A. 1966. 'Observations with self-embedded sentences', *Psychonomic Science, 6/10*, 453–454.

Bock, K. and Loebell, H. 1990. 'Framing sentences', *Cognition, Apr., 35/1*, 1–39.

—— Dell, G., Griffin, Z., Chang, F. and Ferreira, V. 1996. 'Structural priming as implicit learning', Paper presented at the meeting of the Psychonomic Society, November 1996.

Bod, R. (1998). *Beyond Grammar: An Experience-Based Theory of Language*. CSLI Publications: Cambridge University Press.

Boeree, C. G. 2000. Psychology: The Beginnings. http://www.ship.edu/~cgboeree/psychbeginnings.html.

Boot, M. 1976. 'Linguistic Data Structure, Reducing Encoding by Hand, and Programming Languages', in A. Jones and R. F. Churchhouse (eds). *The Computer in Literary and Linguistics Studies*, 255–270. Cardiff: The University of Wales Press.

Bower, G. H. and Winzenz, D. 1969. 'Group structure, coding and memory for digit series', *Journal of Experimental Psychology Monograph, 80/2 (Pt. 2)*, 1–17.

Branigan, H., Pickering, B. and Stewart, A. 1997. 'Syntactic priming of ambiguous sentences in language comprehension', *Language and Cognitive Processes*. Unpublished manuscript.

—— Pickering, B., Liversedge, S., Stewart, A. and Urbach, T. 1995. 'Syntactic Priming', *Journal of Psycholinguistic Research, 24/6*, 489–506.

Cantor, J. and Engle, R. W. 1993. 'Working-memory capacity as long-term memory activation: An individual-differences approach', *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*, 1101–1114.

Caplan, D. and Waters, G. 1999. 'Verbal working memory and sentence comprehension', *Behavioral and Brain Sciences, Feb., 22/1*, 77–126.

Charness, N. 1976. 'Memory for chess positions: Resistance to interference', *Journal of Experimental Psychology: Human Learning and Memory, 2*, 641–653.

—— 1979. 'Components of skill in bridge', *Canadian Journal of Psychology, 33*, 1–6.

Chase, W. G. and Ericsson, K. A. 1982. 'Skill and working memory', in G. H. Bower (ed.). *The Psychology of Learning and Motivation, 16*, 1–58. New York: Academic Press.

—— and Simon, H. A. 1973. 'The mind's eye in chess', in W. G. Chase (ed.). *Visual Information Processing*, 215–281. New York: Academic Press.

Cheung, H. and Kemper, S. 1992. 'Competing complexity metrics and adults' production of complex sentences', *Applied Psycholinguistics, 13/1*, 53–76.

Chi, M. T. H. 1976. 'Short-term memory limitations in children. Capacity or processing deficits?', *Memory and Cognition, 4/5*, 559–572.

Chomsky, C. 1969. *The Acquisition of Syntax in Children from 5 to 10*. Cambridge, MA: MIT Press.

Chomsky, N. 1956. 'Three models for the description of language', *IRE Transactions of Information Theory, IT-2*, 113–124.

—— 1957. *Syntactic Structures*. Mouton: The Hague.

—— 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

—— 1966. *Cartesian Linguistics: A Chapter in the History of Rationalist Thought*, New York: Harper & Row.

—— 1967. 'The formal nature of language', in E. H. Lenneberg (ed.). *Biological Foundations of Language*, 397–442. New York: Wiley & Sons.

—— 1980. 'Comments', in M. Piatelli-Palmarini (ed.). *Language and Learning*. London: Routledge and Kegan Paul.

—— 1995. *The Minimalist Program*. Cambridge, MA: MIT Press.

Christiansen, M. H. and Chater, N. 1999. 'Toward a connectionist model of recursion in human linguistic performance', *Cognitive Science*.

Christidis, A. 1985. 'On generativity: Theory and pre-theory in transformational generative grammar', *Lingua, 66,* 201–224.

Conrad, R. and Hull, A. 1964. 'Information, acoustic confusion and memory span', *British Journal of Psychology, 55/4*, 429–432.

Costa-Pereira, D. J. and Maskill, R. 1983. 'Structure and process in pupils' essays: A graphical analysis of the organisation of extended prose', *British Journal of Educational-Psychology, 53/1*, 100–106.

Cowan, N., Day, L., Saults, J. S., Kellar, T. A., Johnson, T. and Flores, L. 1992. 'The role of verbal output time in the effects of word-length on immediate memory', *Journal of Memory and Language, 31*, 1–17.

Craik, F. I. and Lockhart, R. S. 1972. 'Levels of processing: A framework for memory research', *Journal of Verbal Learning and Verbal Behaviour, 11*, 671–684.

—— and Tulving, E. 1975. 'Depth of processing and the retention of words in episodic memory', *Journal of Experimental Psychology: General, Sep., 104/3*, 268–294.

Cromer, W. 1970. 'The difference model: a new explanation for some reading difficulties', *Journal of Educational Psychology, 61/6*, 471–483.

Crutcher, R. J. 1990. 'The role of mediation in knowledge acquisition and retention: Learning foreign vocabulary using the keyword method', *Tech Report No. 90-10*. Boulder: University of Colorado, Institute of Cognitive Science.

—— 1992. 'The effects of practice on retrieval of foreign vocabulary learned using the keyword method', dissertation, The University of Colorado at Boulder.

Cuetos, F., Mitchell, D. C. and Corley, M. M. B. 1996. 'Parsing in different languages', in M. Carreiras, J. E. Garcia-Albea *et al.* (ed.). *Language processing in Spanish*, 145–187. Mahwah, NJ, USA: Erlbaum.

Cupples, L. and Holmes, V. M. 1987. 'Reading skill and interpretation of temporary structural ambiguity', *Language and Cognitive Processes, 2*, 179–203.

—— and Holmes, V. M. 1992. 'Evidence for a difference in syntactic knowledge between skilled and less skilled adult readers', *Journal of Psycholinguistic Research, 21/4*, 249–275.

Dabrowska, E. 1997. 'The LAD goes to school: A cautionary tale for nativists', *Linguistics, 35/4*, 735–766.

Dailey, C. A. 1959. 'Graph theory in the analysis of personal documents', *Human-Relations, 12*, 65–75.

Daneman, M. and Carpenter, P. A. 1980. 'Individual differences in working memory and reading', *Journal of Verbal Learning and Verbal Behavior, 19*, 450–466.

Day, R. S. 1969. 'Temporal order judgements in speech: Are individuals language-bound or stimulus-bound?', Paper presented at the 9th meeting of the Psychonomic Society, St Louis.
—— 1979. 'Verbal Fluency and the Language-Bound Effect', in C. J. Fillmore, D. Kempler and W. S. Wang, *Individual Differences In Language Ability and Language Behavior*, 57–84. New York: Academic Press.
Deese, J. and Kaufman, R. A. 1957. 'Serial effects in recall of unorganised and sequentially organised verbal material', *Journal of Experimental Psychology, 54*, 180–187.
De Fazio, V. J. 1973. 'Field articulation differences in language abilities', *Journal of Personality and Social Psychology, 25/3*, 351–356.
de Groot, A. 1978. *Thought and Choice and Chess*. Mouton: The Hague. (Original work published 1946).
de Saussure, F. 1916/1990. *Course in General Linguistics*. London: Duckworth.
Devlin, K. 2000. *The Maths Gene: Why Everyone Has It But Most People Don't Use It*. Guernsey: Guernsey Press.
Ellis, R. and Humphreys, G. 1999. *Connectionist Psychology*. Hove: Psychology Press.
Elman, J. L. 1990. 'Finding Structure in Time', *Cognitive Science*, *14/2*, 179–211.
—— 1992. 'Grammatical Structure and distributed Representations', in S. Davis (ed.). *Connectionism: Theory and Practice*. Oxford: OUP.
Engle, R. W. and Bukstel, L. H. 1978. 'Memory processes among bridge players of differing expertise', *American Journal of Psychology, 91*, 673–689.
Epstein, W. 1961a. 'The influence of syntactical structure on learning', *American Journal of Psychology, 74*, 80–85.
—— 1961b. 'A Further Study of the influence of syntactical structure on learning', *American Journal of Psychology, 75*, 121–126.
Ericsson, K. A. and Chase, W. G. 1992. 'Exceptional memory', *American Scientist, 70*, 607–615.
—— and Kintsch, W. 1995. 'Long-term working memory', *Psychological Review, 102*, 211–245.
Eysenck, M. W. 1979. 'Depth, elaboration and distinctiveness', in L. S. Cermak and F. I. M. Craik (eds). *Levels of processing in human memory*. Hillsdale, NJ: Erlbaum.
—— and Keane, M. T. 1995. *Cognitive Psychology: A Student's Handbook*. Hove: Psychology Press.
Ferguson, C. A. 1979. 'Phonology as an Individual Access System: Some Data from Language Acquisition', in C. J. Fillmore, D. Kempler and W. S. Wang. *Individual Differences In Language Ability and Language Behavior*, 189–202. New York: Academic Press.
Fillenbaum, S. 1971. 'Psycholinguistics', *Annual Review of Psychology*, *22*, 251–308.
Fitts, P. M. and Switzer, G. 1962. 'Cognitive aspects of information processing: The familiarity of S-R sets and sub-sets', *Journal of Experimental Psychology, 63*, 321–329.
Fletcher, C. R. 1994. 'Levels of representation in memory for discourse', in M. A. Gernsbacher *et al*. (eds). *Handbook of psycholinguistics*, 589–607. San Diego, CA, USA: Academic Press.
Fodor, J. A. 1983. *Modularity of Mind*. Cambridge, MA: MIT.
—— and Bever, T. G. 1965. 'The psychological reality of linguistic segments', *Journal of Verbal Learning and Verbal Behaviour, 4*, 414–420.
—— and Pylyshyn, Z. W. 1988. 'Connectionism and cognitive architecture: a critical analysis', *Cognition, 28/1–2*, 3–71.

—— Bever, T. G. and Garret, M. F. 1974. *The psychology of language*. New York: McGraw Hill.

Forney, G. D. 1973. 'The Viterbi Algorithm', *Proceedings of the IEEE, 61/3*, 268.

Frase, L. T. 1969. Depth of processing text: Application of graph theory to recall changes induced by the logical analysis of text', *Proceedings of the Annual convention of the American Psychological Association, 4/2*, 617–618.

Frazier, L. 1987. 'Sentence Processing: A tutorial review', in M. Coltheart (ed.). *Attention and Performance XII: The psychology of reading*. London: Laurence Erlbaum Associates.

—— 1995. 'Constraint satisfaction as a theory of sentence processing', *Journal of Psycholinguistic Research 24/6*, 437–467.

—— Taft, L., Clifton, C., Roeper, T. and Ehrlich 1984. 'Parallel structure: A source of facilitation in sentence comprehension', *Memory and Cognition 1984, Sep., 12/5*, 421–430.

Freedle, R. and Craun, M. 1970. 'Observations with self-embedded sentences using written aids', *Perception and Psychophysics, 7/4*, 247–249.

Garret, M., Bever, T. and Fodor, J. 1966. 'The active use of grammar in speech perception', *Perception and Psychophysics, 1*, 30–32.

Garside, R. 1987. 'The CLAWS word-tagging system', in R. Garside, G. Leech and G. Sampson (eds), *The Computational Analysis of English: A Corpus-based Approach*. London: Longman.

Geer, S. E., Gleitman, H. and Gleitman, L. 1971. 'Paraphrasing and remembering compound words', *Journal of Verbal Learning and Verbal Behaviour, 11*, 348–355.

Gernsbacher, M. and Faust, M. 1991. 'The mechanism of suppression: A component of general comprehension skill', *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17/2*, 245–262.

—— Varner, K. R. and Faust, M. E. 1990. 'Individual differences in general comprehension skills', *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 430–445.

Gibson, E. and Thomas, J. 1999. 'Memory Limitations and Structural Forgetting: The Perception of Complex Ungrammatical Sentences as Grammatical', *Language and Cognitive Processes, 14/3*, 225–248.

Gleitman, L. R. and Gleitman, H. 1970. *Phrase and Paraphrase*. New York: W. W. Norton.

Goldberg, A. 1995. *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: University of Chicago Press.

—— 1996. 'Construction Grammar', in K. Brown and J. Miller (eds). *Concise Encyclopedia of Syntactic Theories*. Cambridge: Pergamon.

—— 1999. 'Making One's Way Through The Data', in M. Shibatani and S. Thompson (eds). *Grammatical Constructions: Their Form and Meaning*, 29–53. Oxford.

Goldman-Eisler, F. 1958. 'Speech Analysis and Mental Processes', *Language and Speech, 1*, 59–75.

Graesser, A. C. and Mandler, G. 1975. 'Recognition memory for the meaning and surface structure of sentences', *Journal of Experimental Psychology: Human Learning and Memory, May, 1/3*, 238–248.

Graf, R. and Torrey, A. W. 1966. 'Perception of phrase structure in written language', *Proceedings of the Annual Convention of the American Psychological Association*, 83–84.

Granath, S. 2001. 'Is that a fact? A corpus study of the syntax and semantics of "the fact that" ', in P. Rayson, A. Wilson, T. McEnery, A. Hardie and S. Khoja (eds). *UCREL Technical Papers Special Issue: Proceedings of the Corpus Linguistics 2001 Conference*, 234–244.

Greeno, J. G. 1976. 'Psychological representation of structured knowledge', *Journal of Structural Learning, 5/1–2*, 73–95.

Gross, M. 1979. 'On the failure of generative grammar', *Language, 55/4*, 859–885.

Halliday, M. 1985. *An Introduction to Functional Grammar*. London: Arnold.

Harris, Z. S. 1951. *Methods in Structural Linguistics*. Chicago: University of Chicago Press. Cited in P. H. Matthews. 1993. *Grammatical Theory in the United States from Bloomfield to Chomsky*. Cambridge: CUP.

Heydel, M. and Murray, W. S. 1997. 'Conceptual form and the basis of sentence priming: Cross-linguistic evidence', Poster presentation to *the CUNY Sentence Processing Conference*. Santa Monica, CA.

Hick, W. E. 1952. 'On the rate of gain of information', *Quarterly Journal of Experimental Psychology, 4*, 11–26.

Hockett, C. F. 1948. 'A note on "structure" ', *International Journal of American Linguistics, 14*, 269–271. Cited in P. H. Matthews. 1993. *Grammatical Theory in the United States from Bloomfield to Chomsky*. Cambridge: CUP.

—— 1952. 'A formal statement of morphemic analysis', *Studies in Linguistics, 10*, 210–231. Cited in P. H. Matthews. 1993. *Grammatical Theory in the United States from Bloomfield to Chomsky*. Cambridge: CUP.

—— 1953. 'Review of the Mathematical Theory of Communication by Claude L. Shannon and Warren Weaver', *Language, 29/1*, 69–93.

—— 1954. 'Two models of grammatical description', *Word, 10*, 210–231. Cited in P. H. Matthews. 1993. *Grammatical Theory in the United States from Bloomfield to Chomsky*. Cambridge: CUP.

—— 1955. 'A manual of phonology', *International Journal of American Linguistics, 21/4, Memoir no. 11.*

Holmes, V. M. and Forster, K. I. 1970. 'Detection of extraneous signals during sentence recognition', *Perception and Psychophysics, 7/5*, 297–301.

Hubey, H. M. 1999. *Mathematical foundations of linguistics*. Lincom Europa: Muenchen/NewCastle.

Hudson, R. 1984. *Word Grammar*. Oxford: Blackwell.

—— 1996. *The difficulty of (so-called) self-embedded structures*.

—— 2000. 'Language as a cognitive network', in H. G. Simonsen and R. T. Endresen (eds). *A Cognitive Approach to the Verb. Morphological and Constructional Perspectives*, 49–72. Berlin, Mouton de Gruyter.

Hulme, C., Maugham, S. and Brown, G. 1991. 'Memory for familiar and unfamiliar words. Evidence for a long-term memory contribution to short-term memory span', *Journal of Memory and Language, 30/6*, 685–701.

Hyde, T. S. and Jenkins, J. J. 1973. 'Recall for words as a function of semantic, graphic, and syntactic orienting tasks', *Journal of Verbal Learning and Verbal Behavior, 12*, 471–480.

Hyman, R. 1953. 'Stimulus information as a determinant of reaction time', *Journal of Experimental Psychology, 45*, 188–196.

Imae, K. and Takeuchi, Y. 1959. 'Perception and recall of the series of approximation to Japanese', *Japanese Journal of Psychology, 30*, 168–177. Abstract in PsychINFO.

Jarvella, R. J. 1971. 'Syntactic processing of connected speech', *Journal of Verbal Learning and Verbal Behaviour, 10/4*, 409–416.

Johnson, F. 1965. 'The psychological reality of phrase structure rules', *Journal of Verbal Learning and Verbal Behaviour, 4*, 469–475.

Just, M. A. and Carpenter, P. A. 1992. 'A capacity theory of comprehension: Individual differences in working memory', *Psychological Review, 99*, 122–149.

Karanth, P., Kudva, A. and Vijayan, A. 1996. 'Literacy and linguistic awareness', in B. de Gelder and J. Morais (eds). *Speech and Reading*, 303–316. London: Lawrence Erlbaum.

Keenan, J. M., MacWhinney, B. and Mayhew, D. 1977. 'Pragmatics in memory: A study of natural conversation', *Journal of Verbal Learning and Verbal Behavior, 16*, 549–560.

Kessel, F. S. 1970. 'The role of syntax in children's comprehension from ages six to twelve', *Monographs of the Society for Research in Child Development, Sep., 35/6*, 95.

Kintsch, W. and Bates, E. 1977. 'Recognition memory for statements from a classroom lecture', *Journal of Experimental Psychology: Human Learning and Memory, 3*, 150–159.

Kiss, G. R. 1968. 'Words, associations, and networks', *Journal of Verbal Learning and Verbal Behavior, 7/4*, 707–713.

Kramer, P. E., Koff, E. and Luria, Z. 1972. 'The development of competence in an exceptional language structure in older children and young adults', *Child Development, 43*, 121–130.

Kruijff, G. 2002. Re: Canonical Order. Corpora Mailing List http://nora.hd.uib.no/corpora/2002-4/0212.html (as on 1/4/2003).

Labov, W. 1979. 'Locating the frontier between social and psychological factors in linguistic variation', in C. Fillmore, D. Kempler and W. Wang (eds). *Individual Differences in Language Ability and Language Behavior*, 327–340. New York: Academic Press.

Lachman, R., Lachman, J. L. and Butterfield, E. C. 1979. *Cognitive psychology and information processing: An introduction*. Hillsdale, NJ: Erlbaum.

Lamb, S. 1966. *Outline of Stratificational Grammar*. Washington, DC: Georgetown University Press.

—— 1999. *Pathways of the Brain*: *The Neurocognitive Basis of Language*. Amsterdam and Philadelphia: John Benjamins Publishing Company.

Lane, D. M. and Robertson, L. 1979. 'The generality of the levels of processing hypothesis: An application to memory for chess positions', *Memory and Cognition, 7/4*, 253–256.

Langacker, R. 1987. *Foundations of Cognitive Grammar, I. Theoretical Prerequisites*. Stanford: Stanford University Press.

Le-Blanc, R. S., Muise, J. G. and Gerard, J. 1971. 'Letter reading as a function of approximation to English and French', *Perceptual and Motor Skills, 33/3 (Pt. 2)*, 1139–1142.

Lefton, L. A., Spragins, A. B. and Byrnes, J. 1973. 'English orthography: Relation to reading experience', *Bulletin of the Psychonomic Society, Nov., 2/5-A*, 281–282.

Levelt, W. J. M. 1970. 'Hierarchical chunking in sentence processing', *Perception and Psychophysics, 8*, 99–103.

—— 1974. *Formal grammars in linguistics and psycholinguistics, Vol III: Psycholinguistic applications*. Mouton: The Hague.

Levin, B. and Rapoport, T. 1988. 'Lexical subordination', *Chicago Linguistic Society, 24/1*, 275–289. Cited in Goldberg, A. 1999. 'Making One's Way Through The

Data', in M. Shibatani and S. Thompson (eds). *Grammatical Constructions: Their Form and Meaning*, 29–53. Oxford.
Levin, H. and Kaplan, E. L. 1970. 'Grammatical structure and reading', in H. Levin and J. P. Williams (eds). *Basic Studies in Reading*, 119–133. New York: Basic Books.
Logan, G. D. 1988. 'Toward an instance theory of automatisation', *Psychological Review, 95/4*, 492–527.
Loosen, F. 1972. 'Cognitive organastie van zinen in het geheugen', Unpublished PhD dissertation, Louvain. Cited in W. J. M. Levelt 1974. *Formal grammars in linguistics and psycholinguistics. Vol III: Psycholinguistic applications*. Mouton: The Hague.
Lovatt, P., Avons, S. E. and Masterson, J. 2000. 'The Word-length Effect and Disyllabic Words', *The Quarterly Journal of Experimental Psychology, 53A/1*, 1–22.
Lewis, R. L. 1996. 'Interference in short term memory: the magical number two (or three) in sentence processing', *Journal of Psycholinguistic Research, 25/1*, 93–115.
Lindenmayer, A. 1968. 'Mathematical models for cellular interaction in development', Parts I and II. *Journal of Theoretical Biology, 18*, 280–315.
Liversedge, S., Stewart, A. and Urbach, T. 1995. 'Syntactic priming', *Journal of Psycholinguistic Research, 24/6*, 489–506.
MacDonald, M. 1994. 'Probabilistic constraints and syntactic ambiguity resolution', *Language and Cognitive Processes, 9*, 157–201.
—— Just, M. A. and Carpenter, P. A. 1992. 'Working memory constraints on the processing of syntactic ambiguity', *Cognitive Psychology, 24*, 56–98.
MacDonald, M. C. and Christiansen, M. H. 2002. 'Reassessing working memory: A comment on Just and Carpenter (1992) and Waters and Caplan (1996)', *Psychological Review, 109*, 35–54.
Mackworth, J. 1963. 'The duration of the visual image', *Canadian Journal of Psychology, 17/1*, 62–81.
Maclay, H. and Osgood, C. E. 1959. 'Hesitation phenomena in spontaneous English speech', *Word, 15*, 19–44.
Mandelbrot, B. 1982. *The fractal geometry of nature*. Freeman: San Francisco.
Mandler, G. and Mandler, J. M. 1964. 'Serial Position Effects in Sentences', *Journal of Verbal Learning and Verbal Behaviour, 3*, 195–202.
Maratsos, M. P. 1976. *The use of definite and indefinite reference in young children*. London: CUP.
Markov, A. A. 1913. 'Primer statisticheskogo issledovanija nad tekstom 'Evgenija Onegina' illjustrirujuschij svjaz' ispytanij v tsep (An example of statistical study on the text of 'Eugene Onegin' illustrating the linking of events to a chain)', *Izvestija Imp. Akademii nauk, serija VI*, 3, 153–162.
Marks, L. E. 1968. 'Scaling of grammaticalness of self-embedded English sentences', *Journal of Verbal Learning and Verbal Behaviour, 7/5*, 965–967.
—— and Miller, G. A. 1964. 'The role of semantic and syntactic constraints in the memorisation of English sentences', *Journal of Verbal Learning and Verbal Behaviour, 3*, 1–5.
Marks, M. R. and Jack, O. 1952. 'Verbal context and memory span for meaningful material', *American Journal of Psychology, 65*, 298–300.
Masson, M. E. J. and Miller, J. A. 1983. 'Working memory and individual differences in comprehension and memory of text', *Journal of Educational Psychology, 75*, 314–318.
Matthews, P. H. 2001. *A Short History of Structural Linguistics*. Cambridge: CUP.

McClelland, J. L. and Rumelhart, D. E. 1981. 'An interactive activation model of context effects in letter perception: I. An account of basic findings', *Psychological Review, 88/5*, 375–407.

—— and Rumelhart, D. E. (eds). 1986. *Parallel Distributed Processing, 2: Psychological and Biological Models*. Cambridge, MA: MIT Press.

McDaniel, M. A. 1981. 'Syntactic complexity and elaborative processing', *Memory and Cognition, 9/5*, 487–495.

Mehler, J., Bever, T. G. and Carey, P. 1967. 'What we look at when we read', *Perception and Psychophysics, 2/5*, 213–218.

Miikkulainen, R. 1996. 'Subsymbolic case-role analysis of sentences with embedded clauses', *Cognitive Science, 20*, 47–73.

Miller, G. A. 1956. 'The magical number seven, plus or minus two: Some limits of our capacity for processing information', *Psychological Review, 63*, 81–97.

—— 1958. 'Free recall of redundant strings of letters', *Journal of Experimental Psychology, 56*, 485–491.

—— 1962. 'Some psychological studies of grammar', *American Psychologist, 17*, 748–762.

—— and Isard, S. 1963. 'Some perceptual consequences of linguistic rules', *Journal of Verbal Learning and Verbal Behaviour, 2*, 217–228.

—— and Isard, S. 1964. 'Free recall of self-embedded English sentences', *Information and Control, 7*, 292–303.

—— and Selfridge, J. A. 1951. 'Verbal context and the recall of meaningful material', *American Journal of Psychology, 63*, 176–185.

—— Bruner, J. S. and Postman, L. 1954. 'Familiarity of letter sequences and tachiscopic identification', *Journal of General Psychology, 50*, 129–139.

Mills, J. A. and Hemsley, G. D. 1976. 'The effect of level of education on judgements of grammatical acceptability', *Language and Speech, Oct.–Dec., 19/4*, 324–342.

Mitchell, D. 1994. 'Sentence parsing', in M. A. Gernsbacher (ed.). *Handbook of Psycholinguistics*, 375–409, San Diego: Academic Press.

Mitchell, D. C., Cuetos, F., Corley, M. M. B. and Brysbaert, M. 1995. 'Exposure-based models of human parsing: Evidence for the use of coarse-grained (nonlexical) statistical records', *Journal of Psycholinguistic Research, Nov., 24/6*, 469–488.

Moeser, S. D. 1974. 'Memory for meaning and wording in concrete and abstract sentences', *Journal of Verbal Learning and Verbal Behaviour, 13*, 683–697.

Motter, A. E., de Moura, A. P. S., Lai, Y. C. and Dasgupta, P. T. 2002. 'Topology of the conceptual network of language', *Physical Review E, 65*.

Mowbray, G. H. and Rhoades, M. V. 1959. 'On the reduction of choice reaction times with practice', *Quarterly Journal of Experimental Psychology, 11*, 16–23.

Muise, J. G., Leblanc, R. S. and Jeffrey, C. J. 1972. 'Letter reading by English Ss as a function of order of approximation to French and English', *Psychological Reports, 30/2*, 395–398.

Muncer, S. J. and Bever, T. G. 1984. 'Sensitivity to propositional units in good reading', *Journal of Psycholinguistic Research, 13*, 275–279.

Navon, D. 1984. 'Resources – a theoretical soup stone?', *Psychological Review, Apr., 91/2*, 216–234.

Neisser, U. 1967. *Cognitive Psychology*. New York: Appleton-Century-Crofts.

Nooteboom, S., Wijnen, F. and Weerman, F. 2002. 'Storage and computation in the language faculty', *Studies in Theoretical Psycholinguistics*. Dordrecht: Kluwer Academic Publishers.

Norman, G. R., Brooks, L. R. and Allen, S. W. 1989. 'Recall by expert medical practitioners and novices as a record of processing attention', *Journal of Experimental Psychology: Learning, Memory and Cognition, 15*, 1166–1174.

Oakhill, J. 1994. 'Individual differences in children's text comprehension', in Gernsbacher (ed.). *The Handbook of Psycholinguistics*, 821–848. San Diego: Academic Press.

Onishi, S. 1962. 'The recognition of letter sequences with different orders of approximation to the Japanese language: On the eye-voice span', *Japanese Psychological Research, 4/1*, 43–47. Abstract in PsychINFO.

Perlmutter, D. M. 1971. *Deep and Surface Structure Constraints in Syntax*, New York: Holt, Rinehart and Winston.

Perlmutter, N. J. and McDonald, M. C. 1995. 'Individual differences and probabilistic constraints in syntactic ambiguity resolution', *Journal of Memory and Language, 34*, 521–542.

Perfetti, C. 1994. 'Psycholinguistics and reading ability', in Gernsbacher (ed.). *The Handbook of Psycholinguistics*, 849–894. San Diego: Academic Press.

Peterson, L. R. and Peterson, M. J. 1959. 'Short-term retention of individual verbal items', *Quarterly Journal of Experimental Psychology, 58*, 193–198.

Pickering, M. J. and Branigan, H. P. 1998. 'The representation of verbs: Evidence from syntactic priming in language production', *Journal of Memory and Language, Nov., 39/4*, 633–651.

Pierce, J. R. 1980. *An Introduction to Information Theory: Symbols, Signals and Noise*. New York: Dover Publications Inc.

Pollack, I. and Pickett, J. M. 1964. 'Intelligibility of excerpts from fluent speech: auditory vs. structural context', *Journal of Verbal Learning and Verbal Behaviour, 3/1*, 79–84.

Polya, G. 1954. 'Mathematics and plausible reasoning', *Princeton University Press. Proceedings of the Annual Convention of the American Psychological Association, 4 (Pt. 2)*, 617–618.

Powell, A. and Peters, R. G. 1973. 'Semantic clues in comprehension of novel sentences', *Psychological Reports, 32*, 1307–1310.

Prusinkiewicz, P. and Lindenmayer, A. 1996. *The Algorithmic Beauty of Plants*. Heidelberg: Springer-Verlag.

Quentin, H. 1926. *Essais de critique textuelle*. Paris: Librarie, A. Picard. Cited in Zarri, G. P. 1976. 'A computer model for textual criticism', in A. Jones and R. F. Churchhouse (eds). *The Computer in Literary and Linguistics Studies*, 133–155, Cardiff: The University of Wales Press.

Quirk, R., Greenbaum, S., Leech, G. and Svartvik, J. 1972. *A Grammar of Contemporary English*. London: Longman.

Rabiner, L. R. 1989. 'A tutorial on Hidden Markov Models and selected Applications in Speech Recognition', *Proceedings of the IEEE, 77/2*, 257–286.

Reber, A. S. and Anderson, J. R. 1970. 'The perception of clicks in linguistic and nonlinguistic messages', *Perception and Psychophysics, 8/2*, 81–89.

Reich, P. A. 1969. 'The finiteness of natural languages', *Language, 45/4*, 830–843.

Richardson, P. and Voss, J. F. 1960. 'Replication report: Verbal context and the recall of meaningful English material', *Journal of Experimental Psychology, 60*, 417–418.

Roodenrys, S., Hulme, C., Alban, J. *et al*. 1994. 'The effects of word frequency and age of acquisition on short-term memory span', *Memory and Cognition, 22/6*, 695–701.

Rosenberg, S. 1968. 'Association and phrase structure in sentence recall', *Journal of Verbal Learning and Verbal Behaviour, 7/6*, 1077–1081.

Ross, R. N. 1994. 'Parsing concepts: A discovery technique for poetry therapy', in A. Lerner (ed.). *Poetry in the Therapeutic Experience*. MMB Music, Inc, St Louis.

Rumelhart, D. E. and McClelland, J. L. 1986. *Parallel Distributed Processing, 1: Foundations*. Cambridge, MA: MIT Press.

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

Sampson, G. 2000. *Empirical Linguistics*. London: Continuum International.

Sanders, L. J. 1971. 'The comprehension of certain syntactic structures by adults', *Journal of Speech and Hearing Research, 14*, 739–745.

Schnitzer, M. L. 1993. 'Steady as a rock: Does the Steady State Represent Cognitive Fossilisation?', *Journal of Psycholinguistic Research, 22/1*, 1–19.

Schachter, J. 1996. 'Learning and Triggering in L2 Language Acquisition', in G. Brown, K. Malmkjaer and J. Williams (eds). *Performance and Competence in Second Language Acquisition*, 70–78. Cambridge: CUP.

Scheerer-Neumann, G., Ahola, H., Koenig, U. and Reckermann, U. 1978. 'The use of oral redundancy with reading disabled children', *Zeitschrift fuer Entwicklungspsychologie und Paedagogische Psychologie, 10/1*, 35–48. Abstract in PsychINFO.

Schiffrin, R. M. and Schneider, W. 1977. 'Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory', *Psychological Review, 84*, 127–190.

Scott, R. and Mills, J. 1973. 'Validating the Permutational Test of Grammaticality', *Language and Speech, 16/2*, 110–122.

Service, E. 1998. 'The effect of word-length on immediate recall depends on phonological complexity, not articulatory duration', *Quarterly Journal of Experimental Psychology, 51A/2*, 283–304.

Shannon, C. E. and Weaver, W. 1949. *The Mathematical Theory of Communication*. Urbana: The University of Illinois Press.

Sharman, R. A. 1989. 'An Introduction to the Theory of Language Models', *IBM UKSC Report 204*, July 1989.

Sharp, H. C. 1958. 'The effect of contextual constraint upon recall of verbal passages', *American Journal of Psychology, 71*, 568–572.

Simon, H. A. 1973. 'The structure of ill structured problems', *Artificial Intelligence, 4*, 181–201.

Singh, S. 2000. *The Science of Secrecy*. London: Fourth Estate.

Spencer, N. 1973. 'Differences between linguists and non-linguists in intuitions of grammaticality-acceptability', *Journal of Psycholinguistic Research, 2*, 83–98.

Stevenson, R. J. 1988. 'Memory for referential statements in texts', *Journal of Experimental Psychology: Learning, Memory, and Cognition, Oct., 14/4*, 612–617.

Stolz, W. 1967. 'A study of the ability to decode grammatically novel sentences', *Journal of Verbal Learning and Verbal Behaviour, 6*, 867–873.

St John, M. F. and McClelland, J. L. 1990. 'Learning and applying contextual constraints in sentence comprehension', *Artificial Intelligence, 46/1–2*, 217–258.

Sutton, J. P., Rittenhouse, C. D, Pace-Schott, E. and Stickgold, R. 1994. 'A new approach to dream bizarreness: Graphing continuity and discontinuity of visual attention in narrative reports', *Consciousness and Cognition: An International Journal, 3/1*, 61–88.

Townsend, D. J. and Bever, T. G. 2001. *Sentence Comprehension: The Integration of Habits and Rules*. Cambridge, MA: MIT Press.

—— and Saltz, E. 1972. 'Phrases vs. Meaning in the immediate recall of sentences', *Psychonomic Science, 9*, 381–384.

Traul, G. N. and Black, J. W. 1965. 'The effect of context on aural perception of words', *Journal of Speech and Hearing Research, 8/4*, 363–369.

Waugh, N. C. and Norman, D. A. 1965. 'Primary memory', *Psychological Review, 72*, 89–104.

Werkerly, J. and Elman, J. 1992. 'A PDP approach to processing centre-embedded sentences', in *Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Erlbaum.

Wickelgren, W. A. 1965. 'Short-term memory for phonemically similar lists', *American Journal of Psychology, 78*, 567–574.

Wiles, J. and Elman, J. 1995. 'Learning to count without a counter: A case study of the dynamics and activation landscapes in recurrent networks'. Talk. <http://www.crl.uscd.edu>.

Williams, J. 1995. 'The Psychology of Language', Unpublished Manuscript. Research Centre for English and Applied Linguistics, University of Cambridge, UK.

Wilks, C. 1999. 'Untangling Word Webs: Graph Theory Approaches to L2 lexicons', Unpublished PhD dissertation, University of Swansea.

Wilson, R. J. and Watkins, J. J. 1990. *Graphs: An Introductory Approach*. New York: Wiley.

Wiseman, S. and Tulving, E. 1976. 'Encoding specificity: Relation between recall superiority and recognition failure', *Journal of Experimental Psychology: Human Learning and Memory, July, 2/4*, 349–361.

Wray, A. 2002. *Formulaic Language and the Lexicon*. Cambridge University Press.

Zarri, G. P. 1976. 'A Computer Model for Textual Criticism?', in A. Jones and R. F. Churchhouse (eds). *The Computer in Literary and Linguistics Studies*, 133–155. Cardiff: The University of Wales Press.

Appendix

Example sentence and questions

Test Sentence:	Tom thinks that the fact that keeping clothes clean is absolutely necessary surprises the butcher.
Question 1:	What does Tom know? The fact that keeping clothes clean is absolutely necessary surprises the butcher.
Question 2:	What surprises the butcher? The fact that keeping clothes clean is absolutely necessary.

List A

1. Tom thinks that the fact that keeping clothes clean is absolutely necessary surprises the butcher.
2. Alison believes that the knowledge that finding money quickly is not easy distresses the teenager.
3. Jane claims that the suggestion that leaving children behind is very cruel annoys the man.
4. Bill says that the fact that singing songs loudly is not allowed angers the choir.
5. James thinks that the idea that passing exams well is not important concerns the teacher.
6. Phil believes that the fact that flying planes low is very dangerous surprises the pilot.
7. Janet thinks that the suggestion that keeping babies warm is not important amazes the nanny.
8. Bill announces that the law that reading foreign books is now illegal angers the librarian.
9. Greg says that the fact that stealing clothes openly is really stupid amuses the boy.
10. Harry thinks that the idea that driving every day is too expensive worries the nurse.

List B

1. Heather believes that the suggestion that keeping bodies active is very healthy surprises the girl.
2. Jenny says that the idea that having chocolate everyday is quite unhealthy upsets the secretary.
3. Jill suggests that the comment that waking children early is not bad bothers the doctor.
4. Paul says that the belief that walking dogs often is really necessary irritates the child.

5. Sally writes that the fact that paying rent late is not tolerated worries the student.
6. Arthur says that the suggestion that arriving home late is now permitted pleases the youth.
7. Linda says that the fact that riding bicycles fast is now forbidden annoys the tourist.
8. Grace thinks that the idea that watching television every day is very exciting disgusts the man.
9. Kelly believes that the suggestion that visiting the children often is quite possible reassures the parents.
10. Tom thinks that the fact that keeping offices tidy is quite essential surprises the manager.

List C

1. Harry claims that the fact that changing trains often is really annoying bothers the inspector.
2. Gavin says that the comment that growing vegetables organically is too difficult amuses the farmer.
3. Cory claims that the notion that eating food slowly is really posh irritates the woman.
4. Ron believes that the idea that growing GM foods is really dangerous angers the government.
5. Janet believes that the suggestion that watching movies often is very good amazes the psychologist.
6. Alistair says that the fact that using paper carelessly is very wasteful surprises the pupils.
7. Gail reports that the idea that phoning every day is not allowed worries the mother.
8. Amy claims that the suggestion that making good profits is quite possible encourages the grocer.
9. Ben believes that the notion that drinking beer often is quite normal surprises the doctor.
10. Kim says that the fact that raising children well is very important inspires the couple.

Index