THE PROCESS OF CHANGING REFERENCE IN SIMPLE TEXTS

ALEXANDER WILLIAM ROBB NELSON

DEGREE OF DOCTOR OF PHILOSOPHY
UNIVERSITY OF EDINBURGH
1992

To my parents

Declaration

I declare that this thesis has been composed by myself and that the research reported therein has been conducted by myself with the exceptions described in Appendix E.

Edinburgh, 30 April 1992

Alexander Nelson

Acknowledgments

Thanks to Keith Stenning for making my thesis possible in so many ways. Most importantly, he did this by being my supervisor and by being director of the Human Communication Research Centre which enabled me to combine study and a research job. In both capacities he helped me with ideas and advice.

I am grateful to Rosemary Stevenson for introducing me to psychology and for stimulating my interest in cognitive science. In addition to her teaching abilities I appreciate her use of euphemism which significantly influenced my academic career. I am pleased to have been able to collaborate with her on the work reported in Chapter 3 and to benefit from her further guidance and advice.

Joe Levy collaborated with me on the work reported in Chapter 5. Apart from this contact he has offered much useful advice and assistance on numerous matters, from computer woes to practice as a researcher. His collection of maxims have surely helped me produce this thesis—the most apposite being "Every day is thesis day."

Mukesh Patel, Morag Brown and Garry Wilson have all provided me with data to analyse which is reported in Chapter 2, Chapter 4 and Chapter 3 respectively. I am grateful for their efforts and kindness.

Note

In accordance with regulation 3.5.8 of the Regulatory Standards for Theses I have included a copy of the following paper in Appendix F which contains research reported in Chapter 2.

Stenning, K., Nelson, A. W. R., Levy, J., Patel, M. J. and Gemmell, M. (in press) Representations of individuals and the processing of reference change. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*

Abstract

The thesis investigates the link between memory and the construction of representations of individuals by studying referential change. The issue of how properties of individuals are represented as belonging to one and the same individual in memory is considered, and the processing of simple descriptions of pairs of individuals is considered in relation to this issue.

Previous work has shown that referentially predictable descriptions of individuals impose no cost when attention switched between individuals. This investigation shows that when texts are referentially unpredictable, switching reference does cost time and that the two individuals are treated asymmetrically. Recall analysis supports this and suggests that subjects focus on one of the individuals more than the other. This is explained by assuming that subjects encode simple surface characteristics which allow them to infer the pattern of attribute binding described by a text. When this mapping from surface order to semantics is disrupted by the introduction of unpredictability, subjects respond by treating the individuals asymmetrically which restores the mapping.

It is proposed that this use of surface information is a general feature of language processing. Work on parallel function in pronoun comprehension is a closely related issue and some work had already addressed the issue of surface information effects. Therefore an investigation of parallelism in pronoun comprehension was carried out which revealed effects of surface information along with effects of grammatical parallelism which interacted with animacy and sentence structure.

The investigation of switching reference also revealed word length effects which were interpreted as speech-based memory effects. Further investigations of these effects showed that they were not interfered with by articulatory suppression which contradicts interpretations of similar effects reported in the immediate serial list recall literature. Similar contradictory results were subsequently found in immediate recall tasks using lists with structured vocabulary. These results were interpreted as evidence for the involvement of semantic memory in speech-based memory.

The results of the experiments are discussed and their relevance to discourse processing, pronoun comprehension and working memory is outlined.

Contents

1 General introduction **1**

1.1 Aims of research . . . 1

1.2 Some existing theories of memory . . . 2

1.3 Problems with existing theories . . . 5

1.4 The binding problem and why it matters . . . 6

1.5 Theories which address the binding problem . . . 9

1.6 Extension of the MIT . . . 13

1.7 Working memory . . . 15

1.8 Summary of motivation . . . 16

1.9 Plan of the thesis . . . 16

2 Unpredictable switches of reference in simple texts **18**

2.1 Introduction . . . 18

2.2 Experiment I . . . 22

2.2.1 Method . . . 22

2.2.2 Results . . . 25

2.2.3 Regression modelling . . . 28

2.2.4 Discussion of regression model . . . 32

2.2.5 Switching reference in a text . . . 33

2.2.6 Word length effects and the role of articulatory rehearsal . . . 34

2.3 Experiment II . . . 36

2.3.1 Method . . . 37

2.3.2 Results . . . 38

2.4 General Discussion 43
2.4.1 The purpose of status 43
2.4.2 Implications for memory 44

3 Parallelism in pronoun comprehension **47**

3.1 Introduction 47
3.2 Experiment III 57
3.2.1 Subjects 57
3.2.2 Design and Materials 57
3.2.3 Procedure 58
3.2.4 Results 59
3.2.5 Discussion 62
3.3 Experiment IV 62
3.3.1 Subjects 62
3.3.2 Materials and design 62
3.3.3 Procedure 63
3.3.4 Results 63
3.3.5 Discussion 65
3.4 Experiment V 65
3.4.1 Subjects 65
3.4.2 Materials and Design 65
3.4.3 Procedure 66
3.4.4 Results 66
3.4.5 Discussion 67
3.5 General Discussion of Experiments III, IV and V 69
3.6 Experiment VI 70
3.6.1 Subjects 70
3.6.2 Design and Materials 70
3.6.3 Procedure 71
3.6.4 Results 72
3.6.5 Discussion 74

3.7 Experiment VII 75

3.7.1 Subjects 75

3.7.2 Design and Materials 76

3.7.3 Procedure 77

3.7.4 Results 77

3.7.5 Discussion 78

3.8 Experiment VIII 80

3.8.1 Subjects 80

3.8.2 Materials and Design 80

3.8.3 Procedure 81

3.8.4 Results 81

3.8.5 Discussion 86

3.9 General discussion of Experiments VI, VII and VIII 90

3.9.1 Comparison of Experiments VI and VII 90

3.9.2 Comparison of Experiment VIII with Experiment V 92

3.9.3 Comparison of Experiment VIII with Experiment VI and Experiment VII 92

3.9.4 Summary 93

3.10 General Discussion 94

4 Overt rehearsal in the MIT **97**

4.1 Introduction 97

4.1.1 The role of articulatory rehearsal in the MIT 97

4.1.2 Methodology 98

4.2 Experiment IX 100

4.2.1 Method 100

4.2.2 Results 102

4.3 Discussion 111

5 Articulatory suppression in complex and simple tasks **114**

5.1 Introduction 114

5.2 Experiment X 118

5.2.1 Method 118

5.2.2 Results 120

5.3 Experiment XI 121

5.3.1 Method 121

5.3.2 Results 122

5.3.3 Discussion 124

5.4 Experiment XII 124

5.4.1 Method 124

5.4.2 Results 126

5.4.3 Discussion 129

5.5 General discussion 132

6 General discussion and final remarks **137**

6.1 Summary of results 137

6.2 Conclusions 138

6.3 Future work 140

References **143**

A Appendix to Chapter 2 **155**

A.1 Vocabulary sets used in Experiments I and II 155

A.1.1 Experiment I vocabulary 155

A.1.2 Experiment II vocabulary 155

B Appendix to Chapter 3 **157**

B.1 Materials used in pronoun comprehension experiments 157

B.2 Analyses of question answers in pronoun comprehension experiments 178

C Appendix to Chapter 4 **183**

C.1 The vocabuary set used in Experiment IX 183

D Appendix to Chapter 5 **184**

D.1 Vocabulary used in Experiment X 184

D.2 Results and analysis for Experiment X 184

D.3 Discussion 193

D.3.1 Reading times 193

D.3.2 Recall results 194

D.4 Error-type classification 195

E Details of the author's contribution to the work reported herein 197

F Reprint of paper containing research reported herein 198

List of terms and abbreviations

MIT The Memory for Individuals Task was used to investigate the binding problem by Stenning, Shepherd and Levy (1988), see page 10.

SOE The increase in reading time as more information is discovered about an individual which was first described in Stenning (1986), see page 12.

P×P and I×I Two modes in which texts were presented in Stenning, Shepherd and Levy (1988) described on page 12.

Mode The sequence of reference used in an MIT text, see page 22.

Match and Matchtype If two properties are identical on a particular dimension in an MIT text then the properties are said to match and the dimension is matched. In a text which described two individuals the pattern of matching across dimensions is called the text's match structure or Matchtype. See page 23 for a more detailed description.

Change of reference When the reference of two consecutive sentences is different in a MIT text then reference has changed from the first to the second sentence. (*e.g.*, "The bishop is Dutch" followed by "The dentist is Dutch").

Status An individual in an MIT text can have either *primary* or *secondary* status which refers (roughly) to the degree to which the individual is focussed upon. The status of the two individuals can change or reverse: the individual that was the primary individual can become the secondary individual and the secondary individual can become the primary individual. See page 26 for a full description.

Modsent This is a variable which varies between 1 and 8 in integer steps. The values 1 to 4 correspond to the first four properties of the first individual in an MIT and the values 5 to 8 refer to the first four properties of the second individual. For example, Modsent 6 refers to the second property of the second individual, see page 26.

TOL Thinking-Out-Loud refers to a particular type of experimental data, see page 98.

Rehearsal commentary and segment These terms refer to groups of properties that subjects produce when they externalise their rehearsal in an MIT. See page 104.

Chapter 1

General introduction

Summary

Text processing and memory are intimately linked which means that if text processing is to be understood then a theory of memory is a prerequisite. Models of memory have often relied upon computer metaphors and computer models which has led to the oversight of the problem of how to group properties together to define an individual. The human solution seems to rely on a distributed redundant representation. By considering the construction of such a representation further insight will be gained into its nature which will in turn inform theories of text processing.

1.1 Aims of research

THIS THESIS reports research which was undertaken with the aim of investigating some aspects of human memory underlying text processing. The fact that memory bears a close relation to text processing has been known since the ancient rhetoricians proposed stylistic conventions which were to be adhered to precisely because they produced easily remembered text.

> ... it is easily learned because it is easily memorized, and this is because the periodic style is numerical and number is the most easily remembered thing of all.
> ARISTOTLE, *The Art of Rhetoric*, III. 1409b

In a sense, the properties of memory determine the structure of text because the goal of the text's author is to be understandable and producing a memorable text increases comprehensibility. In addition, when a text is being read, memory must be used to support cohesion. Otherwise relations between sentences might not be so easily extracted.

Modern theories of text processing sometimes invoke aspects of memory as explanations of language phenomena. Miller (1962) proposes an explanation of the unacceptability of centre embedded sentences which depends on a working memory whose capacity is exceeded when there are three or more levels of embedding. Guindon (1985) explains focus phenomena in pronoun comprehension by supposing that potential antecedents are held in working memory whose capacity can be exceeded. When the capacity is exceeded, potential antecedents are no longer available to the resolution machinery so focus phenomena emerge. Clark and Sengul (1979) propose a related theory which similarly supposes working memory capacity limits the number of antecedents which can be held available for pronoun resolution. Berwick and Weinberg (1984) propose a theory of parsing which again depends upon a limited capacity "window" which holds the superficial items to be processed and again this limited capacity is identified with a limited capacity working memory. These examples have all illustrated performance limitations expressed in terms of capacity limitations but as Miller (1956) makes clear in order to understand capacity, the nature of what is represented must be understood.

Memory and representation are considered to be key issues by some theories of text processing which go beyond simple mentions of working memory capacity as performance governors. Such theories will now be described in terms of their representations. All of them assume primitive unanalysed links between concepts which will be shown to be inadequate accounts of a central property of human memory.

1.2 Some existing theories of memory

Tulving (1972) drew a distinction between *episodic* and *semantic* memory. Loosely, episodic memory refers to specific events or episodes and has an autobiographical aspect whereas semantic memory contains knowledge about the world which does not have a autobiographical feel to it. Given this distinction, investigators have found that semantic memory can be consulted very quickly. For example, Smith, Shoben and Rips (1974) showed that people, on average, could decide that a sparrow is a bird in 975 msec., and Loftus and Suppes (1972) showed that people could think of a fruit beginning with letter "P" in 1170 msec. Collins and Quillian (1969, 1970) tried to account for the extraordinary consultation speed by establishing the structure of semantic memory.

Collins and Quillian (1969, 1970) proposed a hierarchical network of concepts as a representation of semantic memory. Each node in the network represented a concept and the links represented the membership relation between sub- and super-ordinate categories. The network was organised on the principal that information is stored as high up in the hierachy as possible in order to minimise the amount of storage. Figure 1.1 shows an example semantic network.

They tested their theory by recording people's verification latencies to statements like "A canary has wings" and argued that the greater the separation of subject and predicate the longer the latency. Their findings supported their hypothesis but several problems were encountered. First, Conrad (1972) found that when subjects were asked to describe a concept, the most frequently mentioned properties also accounted for the latencies which Collins and Quillian had observed which they had not predicted. Second, Rips, Shoben and Smith (1973) showed that the logical organisation embodied in the hierachy was not borne out in their data and further that typicality (for example, a robin is a more typical bird than a chicken) influenced

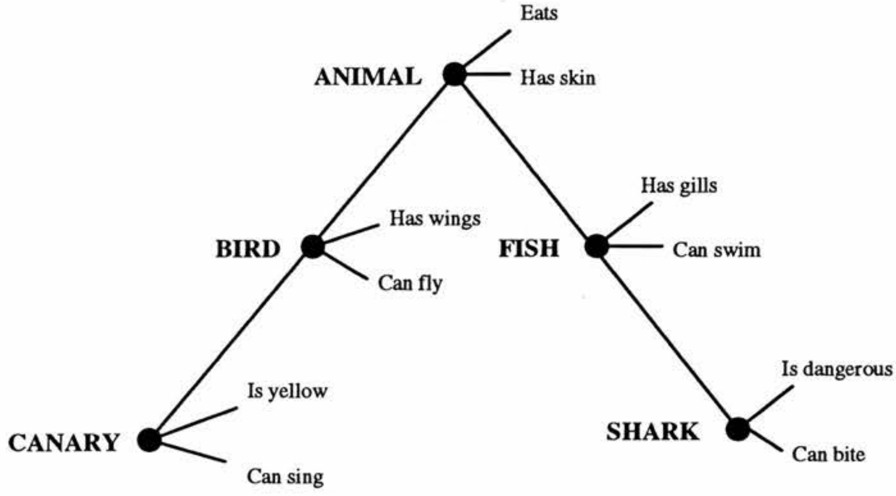

Figure 1.1: A simple hierachical network of the sort proposed by Collins & Qillian (1969).

verification times. Third, Glass and Holyoak (1975) demonstrated that the verification time for false sentences was in fact inversely related to hierarchical distance.

In response to these difficulties Collins and Loftus (1975) proposed a spreading activation model whose structure was based on semantic relatedness. Each node in the network again represented a concept and there were several kinds of links between nodes which might have been "isa" or "isnota" links. Perceptual information activated a node which corresponded to the concept perceived. Activation then passed down the links to other nodes. The spread of activation was determined by the strength of the initial activation, the number of links traversed and the time delay since the perceptual event. There was also a decision process which took account of the nature of the links between concepts and calculated a total activation which was used to decide whether the sentence was true or false.

There are various objections to the theory although it does account for several robust findings, such as, the fast response times for false sentences (there might be a strong "isnota" link) and the fact that semantically related concepts are recognized quickly. One such objection was raised by Ratcliff and McKoon (1981) who showed that activation did not take a significant amount of time to propagate through the network although they were able to support Collins and Loftus' assumption that activation decreases as the number of links traversed increases.

Perhaps one of the most influential theories of memory and cognition is Anderson's (1983) ACT* which is the culmination of a long line of work (FRAN, Anderson, 1972; HAM, Anderson & Bower, 1973; ACT, Anderson, 1976). His general framework assumes three memories: declarative memory, working memory and procedural memory. Declarative memory is a hierarchy of cognitive units which might be propositions (hate, Bill, Fred), strings (one, two, three) or spatial images (a triangle above a square). Figure 1.2 shows an example of a propositional network. The knowledge structures are simply networks made up of nodes, which can have an associated activation, and links which can carry activation. The activation of a node is, in part, a reflection of its frequency of use. Working memory represents the portion of declarative knowledge that is activated and which the procedural component can access. The procedural component is a set of production rules which can match patterns

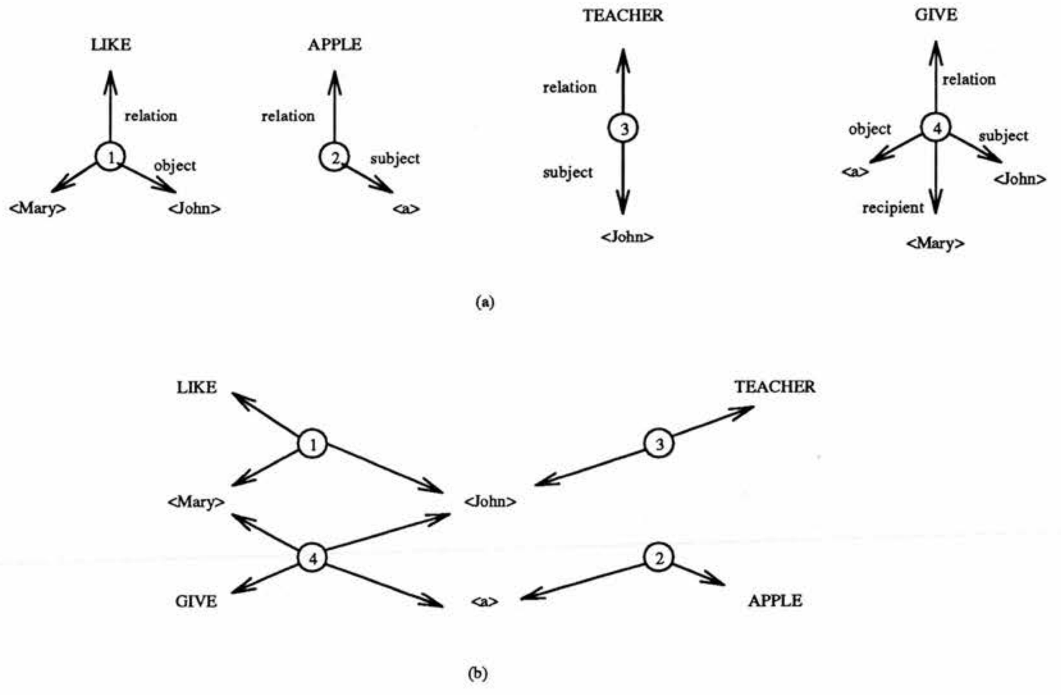

Figure 1.2: Network notation for propositions. (a) Notation for four propositions with all links labeled. (b) The four propositions combined into a single network without the link labels. The propositions are (1) Mary likes John (2) An apple (3) John is a teacher (4) John gives Mary an apple. From Stillings *et al.* (1987), p. 24.

in working memory and produce new propositions, strings or images that can be stored in working memory or in declarative memory. Anderson uses this model to attempt to explain a wide variety of phenomenon such as memory, language, problem solving, imagery, deduction and induction. Here, the most relevant aspect of his theory is knowledge representation.

Kintsch (1980) criticises the research on semantic memory because he claims it is paradigm bound and Johnson-Laird, Herrmann and Chaffin (1984) criticise such theories because they are too powerful (so explaining little) and fail to consider the all important relationship between a semantic network and the external world. Bartlett (1932) proposed an alternative approach which supposed that new material is remembered with reference to known information held in structures called *schemas*. This notion was recouched by Minsky (1975) using the term *frame* and by Schank (1972) using the name *script*. Rumelhart and Norman (1985) characterise these notions by proposing that they all have variables which can be changed depending on the stimulus, they can be embedded, they can represent knowledge at different levels accrued through experience and that the application of a schema will help a person understand what is to be remembered. These ideas have been further developed by Schank (1982) to take account of experimental data collected by Bower, Black and Turner (1979) which showed that people often confused two pieces of information if two schemas were similar. He developed the concepts of *plans*, *scenes*, *memory organization packets* (MOPs) and *thematic organization points* (TOPs) which are used to refine the levels of organization to allow for similarities and increase flexibility.

In spite of these efforts, Kintsch (1989) discussed the use of semantic nets, frames and scripts

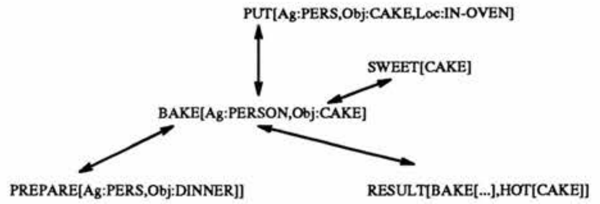

Figure 1.3: Part of an associative net from Kintsch (1988).

sculptor=artist
sculptor=artist
sculptor=artist
(artist)

Figure 1.4: An example of a mental model representation of "All sculptors are artists" from Johnson-Laird (1983).

and concluded that they were too inflexible and could not adapt to the demands of an ever changing context. As an alternative Kintsch (1988) proposed an associative net to represent a minimally organized knowledge system which does not have prestored structure. Nodes in the network represent propositions and links between nodes can have strengths between 1 and −1. Each node consists of a head and a number of slots which hold arguments and specify the relation between the head and the argument. Slots may represent attributes, parts, verb cases or arguments to functions. Figure 1.3 shows part of his proposed associative net which represents the proposition BAKE[agent:PERSON,object:CAKE].

Finally, Johnson-Laird (1983) proposed a theory of mental models which he claimed explained many phenomena in Cognitive Science including aspects of text processing such as text coherence. He proposed two levels of representation: one propositional, possibly held in phonological form in a working memory and a mental model level held in longer term memory. Figure 1.4 illustrates the mental model representation of the sentence "All sculptors are artists". The token **sculptor** represents an individual who is a sculptor and the **(artist)** token represents the uncertain existence of a fourth artist who is not a sculptor. The = token denotes the identity of pairs of individuals. Therefore the token **sculptor=artist** represents one individual who is a sculptor and an artist. Johnson-Laird draws a distinction between mental models and semantic networks by saying that a semantic network is like a proposition and hence describes a set of models whereas a mental model represents that set of models because it is a representative sample of them.

1.3 Problems with existing theories

The main problem with the theories described above relates to general knowledge. Many of them purport to represent general knowledge and aspects of the world which are necessary to comprehend text. Bartlett (1932) characterised this necessity with the phrase *effort after meaning* which meant that a text had to be understood within a body of knowledge or context.

His classic demonstration used a North American Indian folk tale which his subjects were unfamiliar with. They were asked to read the story and then recall it but because they did not have suitable general knowledge to interpret the story they forgot the bits of it which violated their expectations. Bransford, Barclay and Franks (1972) demonstrated that subjects made inferences during comprehension which, of course, depend on general knowledge and that these inferences are stored along with the stimulus material's interpretation. Bransford and Johnson (1972) also demonstrated the importance of a title in setting a context within which to interpret a text. These findings emphasise the importance of the relation between general knowledge and text processing. However, the problem lies not in these aspects of general knowledge but in a more fundamental aspect of contentful memory which Stenning, Shepherd and Levy (1988) have called the *attribute binding problem.*

1.4 The binding problem and why it matters

Essentially the binding problem is how to remember or represent several properties as belonging to one and the same individual. For several individuals the problem can be expressed as a problem of remembering a grouping of properties which defines a set of individuals when there is a multitude of alternatives. For example, a waiter faces this problem frequently when taking orders. From the waiter's point of view the diners may be defined in terms of their orders. For example, *clam chowder*, *steak* and *chocolate gateau* defines one diner and *spinach roulade*, *nut roast* and *fruit salad* defines the other. Clearly the dishes/properties could be grouped or associated in many ways and it is the waiter's task to remember the grouping. Of course, this problem is not limited to restaurants but appears in many types of text. For instance, narrative text introduces individuals and ascribes attributes to them which the reader must surely remember if the text is to be understood; after all, associating villainous properties with virtues in a fairy tale will lead to nonsense.

Many of the foregoing theories of memory have been modelled using computers. One of the reasons that the binding problem has been overlooked is that, for computers, the solution to this problem is trivial. They have a huge array of constructs which allow simple symbolic links to be made between memory locations which suffice to define a particular individual by linking properties to a single "individual" location. One memory location can denote the property *happy* and another location can denote an individual. And so, to ascribe the property of happiness to that individual the memory address of the property *happy* need only be stored at the location of the individual.

The solution for humans may also seem easy. After all, waiters usually get orders right (with a little help from a notepad to cope with the massive interference in an evening) and readers rarely have the experience of becoming so confused about characters in a novel that they derive no understanding of the plot. However, this ease is illusory and demonstrates one of human memory's distinctive qualities. By considering exceptional memory and the methods of improving memory an insight can be gained into how humans solve the binding problem.

One of the most extraordinary memories ever documented belongs to the mnemonist S whose powers were investigated by Luria (1968). S was able to memorise a matrix of 50 digits perfectly after only three minutes study and retain the information up to several years. His ability rested on the use of synesthesia which is the tendency for one sense modality to evoke another. Hence, his comment on Vygotsky's voice, "What a crumbly yellow voice you have" (Luria, 1968, p24). Of course, the use of mnemonics was first established by the bard Simonides (Yates, 1966). He was playing at a banquet at which the guests mocked Castor and Pollux who decided to cause the building to collapse as a punishment. Simonides did not mock the gods who called him away before the disaster thus sparing his life. After the collapse the bodies were so mutilated that they were unidentifiable. Simonides was able to identify their remains by recalling the positions of the guests when he had last seen them. Many mnemonic techniques depend on associating the items to be remembered with a set of already remembered items. For example, the method of loci follows three stages. First, a series of locations (perhaps, rooms in a familiar house) are memorised. Second, mental imagery is used to associate each of the locations with one of the stimulus items. Third, when recall is required, the person simply takes a mental walk through the series of locations recalling the associated item.

Ross and Lawrence (1968) showed that people were able to recall 95% of a list of 40 or 50 items after a single study period using the method of loci. Bower (1973) compared two groups of subjects, one which used the method of loci and the other which did not. Subjects were shown five lists of 20 nouns and the results showed that the mnemonic group recalled 72% of the nouns on average compared to 28% in the non-mnemonic group. There are several other methods which all demonstrate the use of mnemonics: the peg system (Morris & Reid, 1970; Morris & Stevens, 1974), distorting peoples faces as an aid to remembering their name (Morris, Jones & Hampson, 1978), developing a narrative around the test items (Bower & Clark, 1969) and associating a new foreign word with a near-homophonic English phrase (Atkinson & Raugh, 1975).

The techniques described above are all methods of improving memory beyond our untrained ability and may therefore seem to be irrelevant tricks. In fact, they are not really novel because we use similar techniques everyday without being aware of them. Studies of games and non-games have highlighted the use of knowledge as an aid to memory. De Groot (1966) demonstrated that strong chess players were able to reconstruct a chess position with 90% accuracy after a five second exposure compared to weak chess players' performance of 40% accuracy. Similar effects have been found in go and gomoku (Eisenstadt & Kareev, 1975; Rayner, 1958) and in bridge (Charness, 1979; Engle & Bukstel, 1978). Similarly, knowledge of football correlates ($r = 0.81$) with the ability to remember football scores (Morris, Gruneberg, Sykes & Merrick, 1981). Skilled memory has also been observed in non-games domains such as memory for circuit diagrams (Egan & Schwartz, 1979) and memory for computer programs (Shneiderman, 1976 for FORTRAN and McKeithen, Reitman, Reuter & Hirtle, 1981 for ALGOL). These studies demonstrate the "effortless" mnemonics of everyday life because people with greater knowledge do not necessarily have the experience of using their knowledge to improve their performance but that is what they do. In fact, everybody uses general knowledge to remember information about the world and inevitably most people have such a large amount of everyday knowledge that there are few differences between people so its influence is not exposed by casual introspection. When we are required to memorise information from an unfamiliar knowledge domain we must use memory techniques like mnemonics as a substitute for the domain specific knowledge we are used to but which we lack—only the unusual can cope with arrays of random numbers.

The discussion of memory enhancing techniques has strongly suggested that the human solution to the binding problem lies in the pervasive use of general knowledge or at least *content* because domain specific knowledge is used to single out one combination of items from a huge range of possibilities. The reason that the waiter can remember the two orders is because he can simply remember that one is a "healthy" order while the other is not. By remembering these two facts he can pick out the two correct combinations of properties out of the potential multitude. Ericsson and Polson (1988) describe a study of a waiter with exceptional memory for restaurant orders and note that he uses long-term memory encoding to help him retrieve orders. Of course, the associations or "summaries" need not be so easily expressible and the waiter need not have any awareness of the association. Similarly, expert bridge players are able to bind together cards to make up hands by using their superior bridge knowledge to produce associations related to how the cards would be used in a game. Therefore, the human solution to the binding problem depends on the mobilisation of contentful associations between properties to be remembered which is clearly a general property of human memory.

The question of why content is deployed in human memory is still unanswered. However, work on distinctiveness as an explanation for depth of processing (Craik & Lockart, 1972) offers a clue. Moscovitch and Craik (1976) propose that depth of processing effects arise because shallow encodings are inevitably non-distinctive whereas semantic encoding can provide distinct cues for items. As Eysenck and Eysenck (1980) point out, distinctiveness and depth of processing are confounded but they did attempt to separate them and propose that indeed distinctiveness is the key variable: essentially, coding a series of items so that they become less similar will reduce the interference between items thus enhancing recall accuracy.

This recoding of interfering stimuli is exactly the approach taken by Bairaktaris (1990) in his connectionist memory model. Essentially, he offers his associative network a set of patterns which interfere (an encoded alphabet). The network then learns the patterns by associating a random pattern with a very high dimensionality with each stimulus item. Because the associated patterns are random and have a high dimensionality they interfere very little with each other and so they can act as cues for the original patterns. Levy and Bairaktaris (1991) have extended this idea by assuming a general knowledge store provides high dimensional patterns as analogues to the random patterns. A similar process may be taking place in human memory whereby a semantic coding allows knowledge of the world to decrease the interference among a set of items which explains why content is so useful to memory. How content is used is still unclear although Levy and Bairaktaris' suggest that it is used to produce a pattern for association which has a very high dimensionality.

Connectionist models of memory are interesting because they offer an explanation for the need to deploy general knowledge as a way of reducing interference (as described above). They are also interesting because they can capture some of the other definitive characteristics of human memory such as content addressability (McClelland, Rumelhart & Hinton, 1986) and are therefore a promising tool for modelling memory. They are particularly relevant to this introduction because they undermine the assertion that the binding problem is easily solved by computers. As stated above, the binding problem is easily solved by serial computers following Von Neuman's (1987) architecture but it is not easily solved by computers using a parallel architecture with simple computing elements. There are reasons to suppose that parallel distributed processing networks share some of the computational characteristics of human cognition (such as fast pattern recognition and fuzzy matching of items) and so this underlines the fact that humans' ability to solve the binding problem is in fact remarkable and yet amenable to investigation.

A theory of text processing inevitably depends on a theory of memory and indeed memory is often used to explain text processing phenomena. The use of general knowledge in text processing is wide ranging and many models have attempted to describe its use in text processing to account for phenomena like bridging inferences and the constructive aspects of understanding. As described above in the review of theories of memory each model assumes an unanalysed link between nodes, as a primitive. This assumption avoids the binding problem entirely and is the common weakness referred to. Its neglect is understandable because serial computers have no difficulty solving the problem and humans appear similarly competent. However, the effortless binding of attributes in humans is illusory because we are so good at mobilising content to help us that we barely notice the solution. However, when this illusion is recognised the problem becomes significant and when new computer models of memory using connectionist ideas are considered the problem becomes a technological one also. Therefore, the attribute binding problem is no captious subtlety but a problem demanding an account of its human solution which has so far been all but ignored.

1.5 Theories which address the binding problem

Jones (1976) investigated the role played by intrinsic (Jones, 1978, 1979) knowledge in recall. He presented subjects with pictures of an object, painted a particular colour and in a particular location. The position of an item in the sequence of items was also an attribute. After the presentation phase of the experiment subjects were given a written cued recall task where the cues were a particular group of attributes from one of the stimulus items. The hypothesis which Jones tested was called the fragmentation hypothesis which proposed that the memory trace comprised a fragment of the original stimulus. This meant that if a cue was presented which was contained in the fragment then the whole fragment could be recalled. If the cue had not been stored in the fragment then no trace would be recalled. The theory predicted cue symmetry (any cue would be as good as any other) which was observed for all of an object's attributes except for serial position.

Unfortunately, an object's attributes were orthogonal and when this situation does not apply, attributes may interact to form a gestalt which will lead to cue asymmetry (Salzberg, 1976). In spite of that, Jones does show that representations consisting of fragments predict recall better than propositional network models. His investigation is related to the binding problem because his task does impose a version of the problem and he does take seriously the issue of representation of properties. Unfortunately the overlap between two items in his stimulus materials is minimised and only one item at a time is recalled so that his subjects are not really posed a very hard binding problem.

There is a bishop.
There is a dentist.
The bishop is Polish.
The dentist is Swiss.
The bishop is tall.
The dentist is tall.
The bishop is sad.
The dentist is happy.

Table 1.1: An example text from Stenning, Shepherd & Levy (1988).

Stenning, Shepherd and Levy (1988) explicitly identified the binding problem and investigated it by imposing a very hard problem on subjects which does contain a very high overlap between individuals to be remembered. Their intention was to impose a problem which would actually cause subjects to fail even though they were given rich contentful materials and then analyse their errors as a way of inferring the nature of their representations. As has been noted above, humans are very good solvers of the binding problem which means that the nature of the solution is complex and so the binding system must be disrupted in order to view its workings. As Dell (1986) p. 284, observes, " The inner workings of a highly complex system are often revealed by the way in which the system breaks down."

They developed the Memory for Individuals Task (MIT) to impose a difficult binding problem and to be able to study the construction of the resulting representation. Subjects were presented with a series of eight-sentence texts which they read one sentence at a time. Each text described a pair of individuals (which could be a person or an inanimate object), each of which had four attributes drawn from the same four dimensions (profession, nationality, temperament and stature for people and shape, colour, texture and size for objects). The individuals could have identical properties on a dimension or differ on all dimensions except the first dimension which was always the profession for people and shape for objects. Each text therefore made a coherent description and because of the multitude of possible combinations of the presented properties subjects were faced with a severe binding problem which was exacerbated by the possibility of identical properties on a dimension. The binding problem was extreme but the materials allowed rich associations drawn from general knowledge so contentful solutions were available to subjects. Table 1.1 shows an example of the texts that subjects read sentence by sentence.

When subjects were asked to recall the individuals in a text they were offered a menu of items and asked to pick entries which corresponded to the properties ascribed to one of the individuals. They were then re-presented with the menu and asked to describe the other individual. Table 1.2 shows an example of a menu which might appear after reading the text shown in Table 1.1. Stenning, Shepherd and Levy (1988) describe some of the characteristics of the errors which they categorised. Some parts of subjects' recall were highly correlated because some types of errors were coincident whereas other types appeared to be independent. Within a dimension there are strong correlations between individuals and there are strong correlations between some of the properties within an individual's recall. Because of these correlations they proposed that subjects did not represent properties independently but in small groups.

Stenning and Levy (1988) further discussed these groups and described a connectionist model

bishop	dentist
Swiss	Polish
tall	short
happy	sad

Table 1.2: An example recall menu used in the MIT.

of a solution to the binding problem. They considered a solution to the binding problem which uses quantificational facts to bind properties indirectly rather than binding them directly through referential terms. Consider the following set of facts and their possible instantiation:

$(\forall x)(Ax \vee \neg Ax)$	*All people are either French or Polish*
$(\forall x)(Bx \vee \neg Bx)$	*All people are either fat or thin*
$(\forall x)(Cx \vee \neg Cx)$	*All people are either sad or happy*
$(\exists x)(Ax \wedge Bx)$	*There is someone who is French and fat*
$(\exists x)(Bx \wedge Cx)$	*There is someone who is fat and sad*
$(\exists x)(Bx \wedge \neg Cx)$	*There is someone who is fat and happy*
$(\exists x)(Ax \wedge Cx)$	*There is someone who is French and happy*
$(\exists x)(\exists y)(Ax \wedge \neg Ay)$	*Someone is French and someone else is Polish*
$(\forall x)(\forall y)((x \neq y \wedge y \neq z) \rightarrow x = z)$	*There are two people*

From these facts it can be inferred that $(\exists x)(Ax \wedge Bx \wedge \neg Cx)$ (a fat, happy French person) and $(\exists x)(\neg Ax \wedge Bx \wedge Cx)$ (a fat sad Polish person) even although property A is not bound directly to $\neg C$ via referential terms. That is, there is no term which states $(\exists x)(Ax \wedge \neg Cx)$ although such a statement can be inferred. Stenning and Levy (1988)propose that the groups of properties which account for the characteristics of the recall error distribution are quantificational facts of the sort described above. It may seem that a level of regress has been introduced because the quantificational facts appear to need to be bound together also but it is at this level that Stenning and Levy (1988) propose content has its effect. They suppose that associations between properties are recruited which serve to fix or pick out particular combinations of quantificational facts. For example, the association *catholic* might serve to provide a link between Polish and bishop. This idea that remembering more material than necessary appears puzzling until the above discussion of the effects of general knowledge is recalled. By remembering more material, the material become easier to discriminate and hence interferes less.

Given that an indirect solution to the binding problem is adopted by humans the next question is how do they recover the pattern of binding which they were originally exposed to. Stenning and Levy (1988) propose an inference process similar to the one used to recover the two individuals in the example described above. However, their model performs inference in a connectionist manner rather than a symbolic way. The great advantage of this is that the network is able to return an answer when the input is inconsistent and a traditional symbolic inference engine would fail. Of course, when the input is inconsistent the answer that the network produces is not really an inference but the result of the learned constraints applied to the novel input and is to that extent a generalisation. Stenning and Levy (1988) demonstrate that their network can correctly infer the pattern of binding from well formed input and when the input is corrupted, the network produces a very similar pattern of errors to those found in

P×P	I×I
There is a bishop	There is a bishop
There is a dentist	The bishop is Polish
The bishop is Polish	The bishop is tall
The dentist is Swiss	The bishop is sad
The bishop is tall	There is a dentist
The dentist is tall	The dentist is Swiss
The bishop is sad	The dentist is tall
The dentist is happy	The dentist is happy

Table 1.3: A pair of individuals in P×P mode and I×I mode.

the human data. Clearly this model needs to be extended to make clear how content achieves the binding of the quantificational facts and Nelson (1988) describes an attempt using similar experimental materials to develop a connectionist system that models some aspects of subjects' general knowledge about people although there is no account of how the two systems might be linked.

While the MIT is being performed the time taken for subjects to read a text's sentences are recorded and used to investigate the construction processes. Stenning, Shepherd and Levy (1988) refer to Stenning's (1986) finding that reading times are almost wholly determined by the number of properties known of the individual referred to by the current sentence and are virtually unaffected by the other individual's description. Stenning (1986) also found that the reading time increases as more properties are learnt of an individual. Stenning, Shepherd and Levy (1988) refer to these two facts as the *semantic ordinal effect* (SOE hereafter) because the effect depends on reference and the sequence of attributions to an individual. Stenning (1986) was able to conclude that the reference of the current sentence was crucial because he used two contrasting text orderings. In one ordering the reference of consecutive sentences alternated (which was referred to as P×P) and in the other ordering reference continued until the description of one individual was exhausted and the description of the second individual began (referred to as I×I). Table 1.3 shows examples of two such texts. Stenning, Shepherd and Levy (1988) used regression modelling to investigate the contributions of other aspects of a text's structure to the reading times and show that the match structure (the pattern of identical and non-identical properties across dimensions) also predicts reading time which they explain by noting that it affects which intra-individual associations will be useful in discriminating the two individuals.

One of Stenning, Shepherd and Levy's (1988) aims was to rule out articulatory rehearsal as an explanation of the SOE. If subjects were simply repeating an individual's description to themselves then the more properties that were learnt, the longer the time taken to finish repeating them hence a misleading SOE. Stenning *et al.* used two types of vocabulary to construct their texts, one set of words took longer to say than the other set. If articulatory rehearsal was accounting for the SOE then the slope of the increase should have been greater for texts constructed using the longer words. The effect was not found so they concluded the the SOE was genuinely semantic although there was some evidence that there were differences between the two vocabularies at certain points. This effect will discussed below because it is indicative of speech based memory which has interesting properties.

One of the most surprising aspects of Stenning *et al.*'s report was that switching reference had no cost. In P×P modes there was no extra overhead associated with shifting attention to the other individual from sentence to sentence. This result is in contradiction to ACT*'s prediction. Repeated reference to an item in a text will cause the node representing that item to have an increased activation. Activation from this node will propagate down its links and dissipate so that the activation of the node will tend to decrease. The more links that emanate from a node the wider its *fan* is said to be and the faster its activation will decay. When a reference changes to another node the activation of the first node will decay rapidly as a function of its fan and when reference returns to it the time taken to restore activation to a given level will be a function of its rate of decay which is in turn a function of its fan. Therefore, ACT* predicts that changing reference will cost time which will be a function of the complexity of the referent's properties.

As a way of understanding switching reference in the MIT it is useful to view switching reference as shifting attention. It is then clear that switching reference relates closely to the literature on focused attention although much of the work has been perceptual. The crucial link depends on expectation. Because Stenning *et al.* only used two modes of attribution, the pattern of reference was predictable after the second sentence of a text. In the focussed attention literature, Treisman (1964) proposed that expected stimuli are treated differently to unexpected stimuli because the processing system is pre-biased towards them. Posner and Snyder (1975) make a similar point in their discussion of conscious attention and automatic activation. In their view conscious attention speeds performance when a stimulus is expected and slows performance when it is unexpected. If the stimulus is expected then activation can be automatic but if it is unexpected then conscious attention must switch which takes time.

Neely (1977) studied semantic priming in lexical decision to investigate the effects of expectation. The first word presented to a subject was the name of a category which would be followed by a member of a different pre-specified category (*e.g.*, bird followed by the name of a part of building). To investigate expectation he contrasted two conditions. In the first condition the category name was followed by a different, but expected category (*e.g.*, bird-window) and in the second it was followed by a member of the same but unexpected category (*e.g.*, bird-magpie). The theory predicts that decision speed should be slowed in the second condition where the target word was unexpected and conscious attention would need to shift and Neely's results support the prediction. There is further evidence to suggest that switching attention takes a discernible length of time which comes from evoked potential data (Posner, 1978).

Clearly then the focussed attention literature agrees with Stenning *et al.*'s findings. ACT*'s predictions make no mention of expectation or predictability and therefore relate less well.

1.6 Extension of the MIT

Stenning *et al.* studied the construction of representations of individuals using simple predictable texts. Their motivation for using such simple texts was to impose a severe binding problem on subjects (while using contentful materials thus admitting the use of general knowledge) and provide a means of analysing the construction processes used without complicating the experiment by including anaphoric devices (other than definite anaphora), explicit relations between individuals or notions of quantification. Their results clearly demonstrated the worth of their technique because they were able to propose a realistic mechanism for the human solution to the binding problem and describe some aspects of the constructive process which clearly echo other results such as the conscious attention results described above.

As they concede, their materials do not bear a close relation to natural text because they exclude cohesive devices and relations but if they were to use naturalistic texts the binding problem would be less severe and subjects could easily solve it producing perfect recall and therefore uninformative data. Even though the materials are simple they are more realistic that the list-type materials so often used in memory research. Many studies (for examples see Baddeley, 1976) simply present a list of items to be remembered which subjects are then required to recall or recognise in a set of alternatives. Materials in the MIT might appear similar at first glance but a comparison with Wetherick (1975) shows that this is not the case.

Wetherick showed subjects a list of eight words. In one experimental condition the eight words were four pairs of words—the pairs did not necessarily come together in the sequence. Each pair of words came from a semantic category, such as, domestic animals, birds, fishes, parts of the body, colours and girl's names. The average free recall score was 5.23 out of 8. The similarity with the MIT is clear: categories correspond to dimensions and the two items in a pair correspond to the properties ascribed to the two individuals on a dimension. Crucially, there was no analogue of an individual and no repeats as an analogue of matching. Stenning, Shepherd and Levy (1988) found that the mean recall score for texts was 7.45 out of 8. Of course there is no way of statistically comparing these two scores across the two studies (5.23 *vs.* 7.45) but the difference provides compelling support that the MIT's materials aid memory. Presumably the availability of contentful associations and the domain of individuals helps subjects and it is this recruitment of general knowledge that we use in everyday naturalistic situations. Therefore, although the materials may not be naturally occurring they do capture an important naturalistic aspect of "real" text which can be investigated and which has been neglected in much of the literature.

Although the inclusion of cohesive devices and other naturally occurring characteristics of texts would complicate the MIT to an unmanageable degree, the MIT can be extended to investigate more realistic or at least more complicated patterns of reference. This extension is worthwhile for several reasons. First, the referential terms in a narrative text which describe individuals follow particular orderings which do not affect textual cohesion and there is a question what determines these orderings. Second, the predictable text orderings used by Stenning *et al.* appear to relate to theories of attention and this relation could be further investigated by introducing unpredictable orderings. Third, Stenning *et al.* report evidence of speech based memory and investigating different sequences of attribution may shed further light on its use. The next section will elaborate the description of speech based memory.

1.7 Working memory

Baddeley and Hitch (1974) proposed a model of working memory which has been updated by Baddeley (1986) on the basis of later work. The model was originally proposed to explain coding effects in memory tasks and why some tasks interfered and others did not in focused attention experiments. Working memory was intended to replace short-term memory in multi-store theories. Essentially, they propose a unitary store which is composed of slave systems which have different coding characteristics which are controlled by a modality free central executive. The central executive is an attentional mechanism (Baddeley, 1981, p. 22) which has been investigated very little. Of the slave systems, called the articulatory loop and the visuo-spatial scratch pad, the articulatory loop has been the most thoroughly investigated and is the aspect of working memory which is of most interest here.

The articulatory loop was proposed to account for a body of experimental evidence which indicated that some aspects of memory were speech based in immediate serial recall tasks. Items which were phonemically similar reduced recall accuracy but mainly because of transposition errors (Conrad & Hull, 1964). The longer the items in a list take to say the more likely the accuracy will be reduced. However this difference is attenuated if subjects rehearse an irrelevant item like "the" while performing the task (Baddeley, Thomson & Buchanan, 1975). Suppression (the rehearsal of an irrelevant item) also removes the phonemic similarity effect (Murray, 1968) as does unattended speech (irrelevant speech presented in the experimental environment).

To account for these pieces of evidence, they propose a store (called the phonological store) which contains traces of phonemic representations of items and a refreshing process called the articulatory loop. When an item is perceived it is encoded phonemically in the phonological store and decays at a constant rate. For an item to be encoded from a visual stimulus it must pass thought the articulatory process. The articulatory loop can refresh items in the store by running their associated articulatory programs which are directly related to the time taken to say the items. Because all the items in the store decay at a fixed rate the number of items which can be successfully maintained in the store by the articulatory process is determined by the total time to say all of the items which must be less that the time taken for items to decay.

The phonemic confusion effect depends on the nature of the representation of items held in the phonological store. Items which sound similar have similar representations, so when the articulatory process refreshes items in the store it may mistake the order in which items must be refreshed and hence transposition will occur if order is represented over trace activation. When subjects are required to suppress, visually presented items never become encoded in the phonological store because the articulatory process becomes occupied by the irrelevant task. Because the number of stored items depends on the time they take to say, the word length effect is explained. The effect of suppression acts, simply because items cannot be refreshed and so any residual recall depends on a system other than the articulatory loop. Finally, the effect of unattended speech assumes that auditory stimuli have automatic access to the phonological store so that unattended speech reduces recall accuracy because it interferes with the primary stimuli.

The articulatory process may be in some way analogous to the experience of "inner speech" which is a common experience in reading. Therefore, working memory may well be a useful theoretical tool for understanding text processing. Given the slight evidence found in Stenning *et al.* for articulatory effects and the likelihood that they will be enhanced in unpredictable texts (because unexpected shifts of attention may demand extra temporary buffering), working memory is a relevant area of investigation.

1.8 Summary of motivation

The discussion above has motivated the need for an explanation of the attribute binding problem because it is a crucial competence which is difficult for modern (connectionist) theories of memory which take seriously the computational properties of human micro-cognition. The human solution depends on the use of content as shown by the difference in performance between binding problems posed to experts with domain specific knowledge (chess, for example) and the same problems posed to novices. Therefore an account must admit the influence of content. Stenning *et al.*'s theory of binding does precisely this. Their theory is therefore a small step although the larger problem of explaining the interaction between content and binding remains. They also studied the constructive processes through reading time analysis and in doing so studied the effects of referential change in relation to memory. Stenning *et al.* studied predictable orderings of text which fitted well with results from the attention literature on expectation. The next step is to study unpredictable reference change because most texts are not predictable and unpredictability will further test the capacities of the memory systems used to bind properties.

1.9 Plan of the thesis

Given the interesting properties of referential predictability the question arises, what effects does unpredictability have? This is the starting point of the thesis which is discussed in more detail in Chapter 2. In that chapter two experiments using the MIT are described which investigated the effects of unpredictability and produced novel results. The results suggested that subjects exploited a mapping from surface order onto semantic structure as part of their solution to the binding problem. The results also suggested that articulatory rehearsal was being used in the task, just as Baddeley (1986) supposes that the articulatory loop might be used in reading. Therefore, there were two separate strands of investigation which were suggested by these results: the mapping from surface order to semantics and the use of articulatory rehearsal. Although these strands are not unrelated they are different enough methodologically to necessitate a bifurcation of the thesis. Therefore the mapping will be pursued first and then the use of articulatory rehearsal. The relevant literatures are discussed with their strands.

Chapter 3 picks up the notion of a mapping from surface order to semantics and investigates whether serial order is a common basis of representation. This chapter examines the effects of surface order and the interactions these effects have with other aspects of simple texts in the domain of pronoun comprehension. Previous relevant work on pronoun comprehension is discussed here.

The next chapter, Chapter 4, returns to the investigation of the role of articulatory rehearsal in the construction of representations of individuals. An experiment which used the MIT and asked subjects to provide Talking Aloud Protocols of their rehearsal is described and a model of subjects' performance is proposed. Again, relevant literature is discussed in this chapter.

Given the difficulty of diagnosing true articulatory phenomena Chapter 5 uses the technique of articulatory suppression to asses the involvement of articulatory rehearsal in the MIT. The results from this experiment were rather unusual and so a series of experiments were conducted which investigated these anomalies and revealed that the articulatory loop uses, possibly, several codings rather than the single one originally proposed by Baddeley (1986). These findings are discussed in relation to literature on articulatory rehearsal.

In the final chapter, Chapter 6, the various strands of investigation are drawn together and conclusions about the structure of text are drawn. Finally, there is a discussion of future work which would go toward resolving some of the issues highlighted by this investigation.

Chapter 2

Unpredictable switches of reference in simple texts

Summary

Texts usually introduce an individual's attributes in an unpredictable way. When simplified unpredictable texts are presented to people they treat the individuals described by the text in an asymmetric way. They appear to focus on one individual more than the other. The reason they do this is because it allows them to restore a predictable aspect to the text. This aspect is a kind of parallelism which exists in previous experiments which have used predictable texts. By restoring this parallelism subjects are able to use their normal representational machinery which they use in their solution of the binding problem. The study of unpredictable effects also reveals traces of speech-based memory.

2.1 Introduction

As DESCRIBED in Chapter 1 the results from Stenning, Shepherd and Levy's (1988) study were surprising because they conflicted with the predictions from Anderson's ACT* model. His model predicted that switching reference to another individual would cost time because activation would need to be re-established. In contrast Stenning *et al.* found that switching reference predictably did not cost extra time. Perhaps the explanation for this difference is related to activation processes starting earlier in Stenning *et al.*'s predictable texts as experiments on focussed attention suggest. In spite of the difficulty of arriving at a clear explanation for this difference the MIT needed to be extended to unpredictable patterns of reference for several reasons.

First, if predictability was really allowing activation processes to anticipate switches of reference then unpredictability would disrupt this strategy and switching reference would cost time. If it does cost time then it is important for an account of memory to be able to explain how

A.	The ants were eating jelly.
B.	The ants were hungry.
C.	The jelly was grape.
D.	The ants were in the kitchen.
E.	The jelly was on the table.
F.	The kitchen was spotless.
G.	The table was wooden.
H.	The kitchen was equipped with the blender.
I.	The table was against the stove.
J.	The blender was white.
K.	The stove was hot.

Table 2.1: An example of a simple passage from Kieras (1979).

the time cost is related to other factors. Second, although Aristotle (*Rhet.* III. 1409a-b, *lexis eiromene vs. lexis katestrammene.* See also Gotoff, 1979, nn. 65) advised the two predictable orderings used in Stenning *et al.*'s experiment and good texts do indeed have examples of these orderings, many texts do not. Therefore, unpredictable texts must be studied as part of the enterprise of studying normal text processing. Third, many other areas of research study problems which involve unpredictable shifts of reference and so by extending the MIT, more research may become explicable with data from this task. Fourth, if working memory is used during text processing and its capacity is limited then switching reference will reveal the processes involved in moving representations to and from working memory.

There have been several studies which have investigated unpredictable reference but in various different senses. They are related but none have studied the particular sort of unpredictability intended here where the referent of the next sentence will be made as unpredictable as possible while maintaining the same semantic structure investigated in Stenning, Shepherd and Levy (1988).

Kieras (1978, 1981) reports an investigation of coherence in text processing. An example of the texts he studied is shown in Table 2.1. A sentence could cohere with the previous ones if its referent had already been mentioned (*i.e. given* as in Haviland & Clark, 1974) but if the referent was *new* then it could not be integrated until a successive sentence made explicit a link. Therefore, by manipulating the order of sentences in a text, the amount of unintegrated information that had to maintained varied over a text. Kieras argued that the more unintegrated information that needed to be maintained, the greater the processing load.

Clearly the reference of each sentence was unpredictable but so was the nature of each sentence (whether it specified a relation or a property) which could confuse an analysis of switching reference. For example, does Sentence B in Table 2.1 constitute a change of reference, given that the last mentioned referent was "the jelly" or is it a continuation because its subject is the same subject of Sentence A? Additionally, although the number of referents in a text was constant (there were six), subjects did not know what sort of thing there would be nor when they would be introduced and if the relation was specified, if it was indeterminate with which other referent it would be mentioned. Therefore, there were at least, four other types of unpredictability operating aside from unpredictable reference.

Ehrlich and Johnson-Laird (1982) described a similar study which investigated the effects of

Continuous	Discontinuous
The knife is in front of the spoon.	The glass is behind the dish.
The spoon is on the left of the dish.	The knife is in front of the spoon.
The glass is behind the dish.	The spoon is on the left of the glass.

Table 2.2: An example of a referentially continuous and discontinuous text from Ehrlich and Johnson-Laird (1982).

referential discontinuity on discourse cohesion. Table 2.2 shows an example of a referentially continuous and discontinuous text. The continuity of the texts was manipulated by changing the order of the texts as Kieras did. Their texts again showed the similar multiple unpredictability as Kieras' texts do. Their results showed that discontinuity impaired memory and that the reading time for the final sentence of a discontinuous text was long. They interpreted the long time as evidence for the process of integration which had to be postponed until the last sentence of a discontinuous text. Garnham, Oakhill and Johnson-Laird (1982) carried out a similar investigation of referential coherence in stories and found that referential discontinuity led to degraded memory performance. Their explanation depended on the assumption that a disrupted text gave rise to a fragmented representation which was more susceptible to corruption and less amenable to elaboration.

Greeno and Noreen (1974) investigated the effects of different orderings of sentences within a text. Each of their texts specified the same set of concepts which were arranged hierarchically and the effect of different text organizations was to alter subjects' expectations as they processed a text. They found that when subjects expected particular concepts to be mentioned they processed the concepts faster than when they were unexpected. Smith and Foos (1975) were concerned with the process of constructing a representation of linear order. Their study used a set of sentences which described a linear order and they varied the order of the sentences to expose the construction processes. They concluded that introducing new elements which were unrelated to previous ones caused errors because the unrelated items temporarily increased the memory load imposed by the task.

Cirilo (1981) examined the effects of text structure in story comprehension. His main concerns were with local and global coherence phenomena. Local coherence referred to intersentential coreference established through reference matching or bridging inferences. Global coherence depends on some propositions being identified as specially important to a text's global structure which can then cohere through coreference or inference. He investigated a distance effect which affected local coherence and a height effect which operated at the level of global coherence. He assumed a limited capacity short term store which held recently derived propositions. If a new proposition coreferred with an item in short term memory then it would be processed quicker than if the proposition had left short term memory for a longer term store. What determined whether a proposition was in short or long term memory was its distance from the point of its coreference. The height effect simply depends on the assumption that globally important propositions are maintained in a special buffer for high level propositions which means they will be accessed quicker than lower level ones. The results show that both effects can be manipulated under certain circumstances.

The foregoing description shows that when the pattern of reference within a text has been explicitly investigated it has been used as a way of manipulating the processing load imposed by maintaining unresolved information which is intended as an explanation for why incoherent texts are difficult to understand. Cirilo (1981) takes a slightly different view that given information cannot be maintained in short term memory for the entire duration of a text and that it must be transferred to a more permanent store. When an attempt to establish coreference with the short term propositions fails, a search of long term memory ensues which causes the delay in processing. Therefore Cirilo's explanation does not lie in the assumption that remembering unintegrated information costs extra effort but reactivating information, rather like ACT*'s explanation, takes time.

A theory of switching reference will inevitably be part of a theory of anaphora resolution or verbal reasoning. In anaphora and in particular pronoun resolution studies, the aim of the research is to provide a theory of how antecedents for pronouns are selected and how the relevant discourse model/representation is updated. Oakhill, Garnham and Vonk (1989) concentrate on this latter issue and they do consider patterns of reference citing Lesgold, Roth and Curtis (1979) as an investigation of distance between antecedent and pronoun. They do consider memory issues as well but neglect to consider what effects are simply attributable to the pattern of reference in a text in isolation of effects of cohesion. Pronouns refer and so the pattern of pronouns and antecedents in a text will inevitably affect the resolution process. Therefore any complete account of pronoun resolution processing must take account the effects of different patterns of reference. A similar line of thinking applies to verbal reasoning. In verbal reasoning tasks like term series problems the order of the terms is often manipulated. If the effects of these manipulations are to be fully accounted for then the pattern in which the terms refer must be considered (*e.g.* Smith & Foos, 1975). Therefore producing an account of the effects of different patterns of reference in texts should ultimately inform theories of anaphora resolution and verbal reasoning.

Switching reference is an interesting area to study in text processing for many reasons: it is a common occurrence in natural text, it depends on working memory, it is a component of many other more complex processes and it has apparently thrown up conflicting results. Unfortunately simple unpredictable reference switching not associated with memory load or other theories has not been studied. The rest of this chapter will describe two experiments which were designed to investigate the effects of unpredictable reference switching in simple descriptions of individuals using the MIT.

Sentence Position	Individual 1	Individual 2
1	1	
2	2	1
3	2 3	1 2
4	2 3 4	1 2 3
5	2 3 4	1 2 3 4
6	2 3 4	2 3 4
7	3 4	3 4
8	4	4

Table 2.3: Possible sentence positions of property dimensions (1-4) describing each individual

	Sentence Position							
Mode	1	2	3	4	5	6	7	8
1	a 1	a 2	a 3	a 4	b 1	b 2	b 3	b 4
2	a 1	a 2	a 3	b 1	a 4	b 2	b 3	b 4
3	a 1	a 2	b 1	b 2	b 3	b 4	a 3	a 4
4	a 1	b 1	a 2	a 3	b 2	a 4	b 3	b 4
5	a 1	b 1	b 2	a 2	a 3	b 3	a 4	b 4
6	a 1	b 1	b 2	b 3	a 2	a 3	b 4	a 4
7	a 1	b 1	b 2	b 3	b 4	a 2	a 3	a 4

Table 2.4: Sentence sequences in each mode (letter denotes individual and number denotes property).

2.2 Experiment I

2.2.1 Method

Subjects

Nine postgraduate students of Edinburgh University acted as paid (£4) volunteers.

Design and Materials

Subject read 112 texts in a self-paced task. Each text consisted of eight sentences which described two individuals on four dimensions. The four dimensions were the individual's shape, colour, texture and size, so each sentence took the form *The INDIVIDUAL is PROPERTY*, for example, "The pyramid is hot". The individual mentioned in the first sentence was called the first individual and the sequence of references in a text was called a *mode*. There are 35 possible modes and Table 2.3 shows all the possible sentence positions for each of the eight sentences of a text.

In this experiment texts were presented in 7 of the 35 possible modes. The 7 modes were chosen so that each of the eight sentences occurred in all the possible positions they could over the set of 7 modes.

Table 2.4 shows the sequence of sentences in the 7 modes: letters (a and b) denote the two individuals and numbers (1–4) denote the property dimensions in their sequence of occurrence.

Dimension		
1	2	3
+	+	+
+	+	-
+	-	+
+	-	-
-	+	+
-	+	-
-	-	+
-	-	-

Table 2.5: The 8 matchtypes used in Experiment I. The three symbols in a row represent the matching status of the three non-introducing dimensions in temporal order of their attribution to the first individual, where "+" denotes a match and "-" a mismatch.

Mode 4 (forward) +−−	Mode 1 (backward) −+−
There is a cylinder	There is a narrow thing
There is a pyramid	The narrow thing is soft
The cylinder is red	The narrow thing is white
The cylinder is cold	The narrow thing is a block
The pyramid is red	There is a wide thing
The cylinder is thick	The wide thing is hard
The pyramid is hot	The wide thing is white
The pyramid is thin	The wide thing is a beam

Table 2.6: Two example texts in Forward and Backward format in Mode 4 and Mode 1 showing two matchtypes.

Two sequences of property dimensions called *formats* were used. In the *forward* format which was used by Stenning, Shepherd and Levy (1988), property dimensions appeared in the order $\langle shape, colour, texture, size \rangle$. In the *backward* format the order of property dimensions was $\langle size, texture, colour, shape \rangle$. The format was the same for both individuals in a text.

Individuals could have the same property or different properties on a dimension. If the properties were the same (*i.e.* identical) then the dimension was called *matched* and if the properties were different then the dimension was called *mismatched.* The first property ascribed to an individual was called the *introducer* and never matched. The other three property dimensions could match and mismatch in any combination: the particular combination, or match structure was called the *matchtype.* Table 2.5 shows the 8 possible matchtypes. Table 2.6 shows two example texts.

The full design consisted of following within-subject factors: Mode (7 levels), Format (forward *vs.* backward), Individual (2 levels) and Property (4 levels).

Appendix A.1.1 shows the vocabulary used, which contains 48 words, twelve each denoting shape, colour, texture and size. The groupings "texture" and "size" are approximate. Each dimension (shape, colour, texture, size) contains 6 contrastive nouns or adjectives.

Nine lists of materials, each of 112 texts, were generated. Format and Mode were crossed, so there were 8 texts in each of the 14 conditions in a list of materials. Each Matchtype occurred 14 times in each list but it was randomised with regard to Mode and Format. A text was generated by applying a Mode, a Format and a Matchtype to a prototype pair of individuals. A prototype pair of individuals was generated by randomly picking a pair of contrastive nouns or adjectives from each of the dimensions in the vocabulary set. Once the four pairs of properties had been picked then the properties were assigned to individuals after applying the matchtype of the individuals. For example, the pairs of properties (cylinder/pyramid) (red/green) (hot/cold) (thin/thick) and matchtype +–– could be used to produce the two individuals, thin, hot, red pyramid and thick, cold, red cylinder. If a dimension was matched then one of the pair of properties was selected at random and assigned to both individuals. If a dimension was mismatched then the two properties were assigned at random to the two individuals. Once the two individuals had been generated then the Mode and Format could straightforwardly be applied to make the final text.

One subject was assigned to each material list.

Procedure

The material lists were presented to subjects over 8 sessions. During a session subjects read 7 texts in each Format, and in each Mode making 14 texts altogether. The majority of subjects completed the sessions in 2 or 3 sittings.

The task was a self-paced reading task and they were instructed to read the texts as quickly as possible consistent with recalling them accurately. The texts were presented one sentence at a time on the screen of a BBC microcomputer. Each text was preceded by a *setting* which consisted of the pairs of nouns or adjectives used to generate the individuals described in the subsequent text: for example, (cylinder/pyramid) (red/green) (hot/cold) (thin/thick). The setting remained visible until the subject pressed the space bar whereupon the screen was cleared and the first sentence displayed. Subjects were asked to press the space bar as soon as they had read and understood the sentence whereupon the screen cleared and the next sentence was displayed. Once the final sentence had been read the screen was cleared and a simple question was displayed, such as, "Was there a large square?" After the question had been answered by pressing either the Y (for yes) or the N (for no) key the subjects were asked to recall, using a menu selection system, the individuals described by the most recent text.

The subject was cued to recall the individuals in the presented order or the reverse order. A menu was then displayed which offered subjects the choice between the two contrasted properties on each dimension. For example, if the individuals had been generated from the following pairs of properties, (cylinder/pyramid) (red/green) (hot/cold) (thin/thick), then the following menu would have been offered:

cylinder	pyramid
red	green
hot	cold
thin	thick

Subjects recalled an individual by picking out four properties from the menu. After one individual had been recalled then the subjects were asked to recall the other individual. When both individuals had been recalled a single-sentence description of both of the individuals was

	Where criteria agree		Where criteria conflict		
Mode	Frequency	Cued=Bestfit	Frequency	Cued	Bestfit
1	134	7.31	8	4.25	7.31
2	133	7.45	9	4.11	6.78
3	93	7.12	49	2.65	7.10
4	130	7.14	12	4.08	6.33
5	119	7.25	23	3.22	6.91
6	99	7.03	43	3.09	7.33
7	97	7.23	45	2.8	7.07

Table 2.7: Mean Recall scores by mode and by agreement *vs.* conflict of scoring criteria

displayed (*e.g.*, "There was a narrow soft white block and a wide hard white beam") as feedback. Subjects were provided with no other feedback on the accuracy of their recall. Subjects were then asked to press the RETURN key to begin the next trial.

The recall for each text was recorded and the reading time for each sentence was recorded in centiseconds.

2.2.2 Results

Recall

Recall scores were assigned to texts by awarding one point for every correctly recalled property of each individual. The maximum score was therefore 8. Two methods were used to decide the order in which the individuals had been recalled which meant that there were two scores for each text (Stenning, Patel & Levy, 1987, p. 17, contains a discussion of the reasoning behind the two scoring procedures). One method assumed that the subject had recalled the two individuals in the order which was asked for. This score was called the *Cued* score. The other method calculated two scores corresponding to the assumptions that the subject had recalled the individuals in the correct order and in the reverse order. Whichever score was higher was called the *Bestfit* score. Of course, sometimes the Cued score was the same as the Bestfit score and other times the Bestfit score was higher. Table 2.7 shows the mean number of properties correctly recalled out of 8, tabulated by the two scoring methods, Cued and Bestfit and Mode.

There are two sources of evidence in this table. One source is the frequency of instances where the two scoring methods agree compared to the frequency of instances where they disagree. The second source, is the comparison of the relative recall accuracies by the two methods for each mode.

The occasions when the Bestfit score was higher than the Cued score were interpreted as instances where the order of introduction of the individuals had been confused by the subject. In Modes 3, 5, 6 and 7 the individuals are confused about 4 times as often (average number of confusions for Modes 3, 5, 6 and 7 is 40 and the average number of confusions for Modes 1, 2 and 4 is 9.7) as in the other three modes.

In Modes 1, 2 and 4, when the two scores disagree in their assignments, the mean score by

Bestfit is lower than in cases where there is no conflict ($6.66 < 7.30$, $F(1,426) = 11.62, p < .001$). This shows that forgetting of more than just temporal information must have occurred and that the difference between Cued and Bestfit scores cannot be attributed to introduction-order confusion alone. However, in Modes 3, 5, 6 and 7, when conflict arises, the recall score by best-fit is as high as for cases without conflict ($7.13 = 7.16$, $F(1,566) = 0.012, p > 0.1$). The subject is confused about the order of introduction of the individuals but remembers as well as in trials where such confusion does not occur. It appears from these results that the 7 modes fall into two groups of modes. Modes 1, 2 and 4 will be called *Modegroup 1* and Modes 3, 5, 6 and 7 will be called *Modegroup 2*. By the Bestfit criterion there is no significant difference in accuracy between the two groups of modes ($7.26 = 7.15$, $F(1,992) = 2.61, p > 0.1$).

The property of Modegroup 2 modes which seems most likely to be causing the confusions in introduction-order is that readers learn much about the second individual early on in texts. If it is the case that readers are focussing on the first introduced individual for some representational reason, directly related to the individual being first, then it is not surprising that if the second individual's description arrives before much of the first then the second individual is treated like the first. Furthermore if the subjects are relying on identifying the focussed individual with the first introduced individual then it is clear why some modes cause confusions about the order of introduction. The individual which is focussed upon will be said to have *primary* status and the other individual will have *secondary* status.

For the analysis of reading times it was necessary to have a characterization of the primary/secondary status of the individuals at each sentence. In order to separate the two groups of modes according to their patterns of confusions it was assumed that the first introduced individual was the primary individual unless the the second property of the second individual was learnt before the second property of the first individual, in which case the two individuals changed status. For example, in Mode 5 the sequence of the first three sentences in A1, B1, B2 (where A refers to the first individual, B to the second individual and the numbers refer to their properties). When the reader reaches sentence three, A2 has not been encountered so there is a change of status and B2 becomes a sentence about the primary individual. However, Mode 3 followed the same pattern of recall confusions as Modes 5, 6 and 7 so an extra rule was added which stated that there would be a change of status if B1 was preceded by A2 and immediately followed by B2. This extra rule meant that the status of the individuals changed at Sentence 4 of Mode 3.

Reading Times

Figure 2.8 shows mean reading times by mode and *modsent*. Modsent is a variable created to be consistent with Stenning, Shepherd and Levy's (1988) terminology and covers the combinations of individual and property. The first four properties of the first-introduced individual correspond to the first four values of Modsent (1, 2, 3, 4). The second individual's four properties correspond to the last four values of Modsent (5, 6, 7, 8).

An analysis of variance was carried out, with subjects as the random factor, and mode, individual, property, and format as fixed factors. There was a main effect of mode ($F(6,48) = 8.33, p < 0.0001$), (see Table 2.8 for means for each mode). Mode 1 had the fastest mean

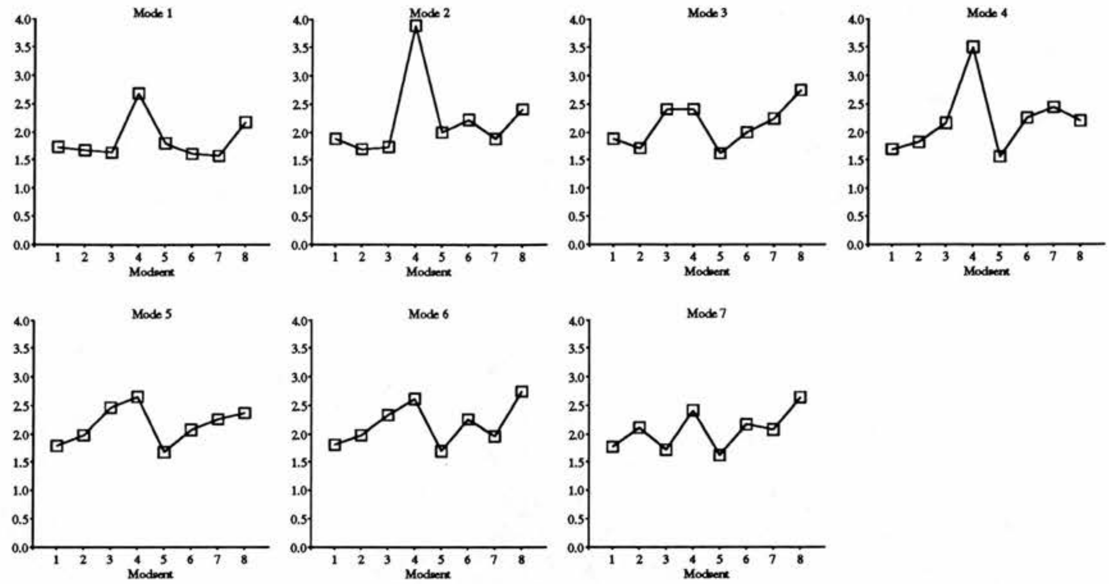

Figure 2.1: Experiment I mean reading times (sec) as a function of Modsent and Mode

	Individual1				Individual2				
Predicate:	1	2	3	4	1	2	3	4	Mean
Mode1	1.72	1.67	1.64	2.68	1.81	1.62	1.58	2.18	1.86
Mode2	1.87	1.69	1.73	3.88	1.99	2.21	1.87	2.41	2.20
Mode3	1.88	1.71	2.40	2.40	1.61	2.00	2.24	2.74	2.12
Mode4	1.70	1.82	2.15	3.51	1.56	2.26	2.43	2.19	2.20
Mode5	1.79	1.97	2.46	2.65	1.67	2.07	2.26	2.37	2.15
Mode6	1.81	1.98	2.32	2.61	1.69	2.25	1.96	2.73	2.17
Mode7	1.76	2.10	1.71	2.41	1.61	2.15	2.06	2.62	2.05
Mean	1.78	1.84	2.06	2.87	1.71	2.08	2.06	2.46	2.11

Table 2.8: Experiment I mean reading times (sec) as a function of individual, property and text mode

reading time per sentence followed by modes 7, 3, 5, 6, 4 and 2 respectively. Figure 2.1 also shows the data graphically.

There was a main effect of property ($F(3,24) = 21.73, p < 0.0001$). This effect was due to a rise in reading time from properties one to four (means were 1.74, 1.96, 2.05 and 2.66 seconds respectively).

There was no main effect of individual ($F(1,8) = 1.56, p = 0.25$). and there was a main effect of format, ($F(1,8) = 16.44, p < 0.004$). Texts in forward format were read faster (mean reading time 1.99 seconds.) than those in backward format (mean reading time 2.21 seconds.). However, there was no significant interaction between format and any other factor.

The interaction between mode and property was significant ($F(18,144) = 5.7, p < 0.0001$), as was the interaction between individual and property ($F(3,24) = 13.06, p < 0.0001$), and the interaction between mode, individual and property factors ($F(18,144) = 5.59, p < 0.0001$). The significance of these last two interactions shows that although the increase in reading time with properties known of the referenced individual appears in this data, it is not completely independent of the temporal sequence of attributions to the two individuals. The obvious explanation is the uneven distribution of processes of referential change in the different modes.

Even from a cursory examination of these results, it is possible to see that they constitute a replication of the general observations of Stenning, Shepherd and Levy (1988). Reading time increases as more is known about the referent and is relatively independent of how much is known about the other individual. The lack of interaction between format and other variables, coupled with the replication of the main effect of number of properties known of the referenced individual, clearly demonstrates that it is the increasing number of known properties, not the dimension from which they are drawn that is controlling this effect. It is the place in the sequence of attributions which determines reading time. Recall error analyses which further support this conclusion are reported in Stenning, Patel and Levy (1987).

2.2.3 Regression modelling

The analysis in the previous section which uses ANOVA to confirm patterns in the reading time data is useful for investigating general properties of referential change. For example the interaction between Mode, Property and Individual shows that the sequence of reference in a text does have a differential effect on the processing of the two individuals. However it is difficult to derive a detailed description of the sentence by sentence processes. Stenning, Shepherd and Levy (1988) used multiple regression techniques to model the construction processes in their simple texts and a similar approach will be adopted here to model these more complicated texts.

Regression modelling is a very sophisticated statistical technique and because of its complexities and power, uninformative models can be derived. Therefore, the approach adopted here insists that new models must replicate and extend previous models and that new variables must have a clear interpretation to guard against adding variables whose only use is to increase the success of the model's predictive powers (see Cohen & Cohen, 1983; Draper & Smith, 1981

Model	Model 1		Model 0	
	$R^2 = 0.123$	DF. = 22	$R^2 = 0.120$	DF. = 11
Variable	Coeff(sec)	St. Error	Coeff(sec)	St. Error
Intercept	1.00		1.09	
FORMAT	0.25	0.03	n/a	n/a
NEUT1	0.29	0.06	0.28	.049
NEUT2	0.30	0.08	0.47	.092
NEUT3	0.45	0.10	0.73	.09
NEUT4	1.18	0.15	2.07	.09
MIS1	0.33	0.06	0.46	.06
MIS2	0.53	0.07	0.79	.08
MIS3	0.53	0.09	1.15	.09
MIS4	0.42	0.14	1.37	.15
LOCALMIS	—	—	0.19	.07
MAT1	—	—	0.42	0.06
MAT2	—	—	0.58	0.09
FORE1	0.18	0.07	—	—
FORE2	0.42	0.11	—	—
FORE3	0.55	0.15	—	—
PRFORE2	0.80	0.18	—	—
PRFORE3	1.10	0.17	—	—
REF+NON-REFSYLL	0.22	0.03	—	—
REFSECSWSYL	-0.24	0.03	—	—
NON-REFSYL	-0.18	0.04	—	—
NON-REFPRSYL	-0.17	0.03	—	—
NON-REFSWSYL	0.17	0.02	—	—
NON-REFSECSWSYL	-0.47	0.04	—	—
REFRUNSYL	-0.04	0.02	—	—
REFPRSYL	-0.16	0.02	—	—

Table 2.9: Summary of reading time Model 1 with primary *vs.* secondary distinctions and syllabic variables. The Stenning, Shepherd and Levy (1988) model (referred to as Model 0) is shown for comparison. The following naming convention is used and described in the text. REF=referenced, NON-REF=non-referenced, PR=primary, SEC=secondary, SW=switched reference, RUN=accumulation since reference switch, SYL=syllables, PROP=properties. For example, NON-REFSECSWSYL refers to the number of syllables in the secondary individual's description when it is the non-referenced individual after a switch of reference.

for guidance on the use of regression modelling). Because of the restriction that models must replicate and extend previous ones, the model described here incorporates Stenning's (1986) observation that reading times are a function of the amount of information known about the currently referenced individual. Table 2.9 shows the model derived by Stenning, Shepherd and Levy (1988) which illustrates the use of variables representing particular types of information. Their model will be referred to as Model 0 hereafter.

The variables NEUT1 to NEUT4, MIS1 to MIS4, MAT1 and MAT2 are dummy variables corresponding to the variables NEUTLOAD, MISLOAD and MATLOAD. Dummy variables are used in multiple regression to allow for non-linear functions (see Draper & Smith, 1981, for a discussion of this use of dummy variables in multiple regression). Briefly, the N levels of a pseudo-continuous variable are represented by N-1 binary variables, each defined so they take the value 1 at their unique level of the parent variable, and otherwise the value 0. There was no prior reason for supposing that these variables might be linearly related to reading

	MIS-LOAD	MAT-LOAD	NEUT-LOAD	CONTOUR	FORE-GROUND	FORMAT
Text			Mode 4 (forward) +−−			
There is a circle	0	0	1	0	0	0
There is a square	1	0	0	1	0	0
The circle is red	1	0	1	1	1	0
The circle is cold	1	0	2	0	0	0
The square is red	2	0	0	1	1	0
The circle is thick	2	0	2	1	3	0
The square is hot	2	1	0	1	2	0
The square is thin	3	1	0	0	0	0
			Mode 7 (forward) −−+			
There is a triangle	0	0	1	0	0	0
There is an oval	1	0	0	1	0	0
The oval is yellow	1	0	1	0	0	0
The oval is dry	1	0	2	0	0	0
The oval is solid	1	0	3	0	0	0
The triangle is blue	1	1	0	1	1	0
The triangle is wet	2	1	0	0	0	0
The triangle is solid	3	1	0	0	0	0

Table 2.10: Two example of values taken by each variable

time so that constraint was not imposed.

NEUTLOAD corresponds to the number of properties of the referenced individual whose matching status is unknown. For example when the subject reads that there the pyramid is hot she may have not known whether the cylinder is hot or cold. If that is the case then the value of NEUTLOAD is increased by 1. MATLOAD corresponds to the number of property dimensions which are known to match across the two individuals for a text and MISLOAD corresponds to the number of property dimensions which are know to mismatch. LOCALMIS is a binary variable which is only ever 1 when a property mismatch is determined by the current sentence—the rest of the time it is 0. Examples of these variables are shown in Table 2.10.

It is clear from the model in Table 2.9 that the function relating reading time to knowledge of the referenced individual is sensitive to relations between the properties of the two individuals described in a text. When a new property is learnt, its match status affects the sentence's reading time because the MISLOAD variable has non-zero coefficients and the MATLOAD variable is absent. This means that if the property mismatches the property on the same dimension of the other individual's description then the sentence will take longer to read than if the property had matched.

The actual process of model building is an iterative one. First a pool of variables are developed which code various aspects of the information available to subjects as they process a text. These variables are designed so that they have a clear interpretation and are ones which are thought to be important on the basis perhaps of other experiments or analyses. A stepwise model-building computer program (BMDP, Dixon, 1988) is then used which selects a set of variables on the basis of their predictive power. The program then reports which variables have been selected and what their coefficients are. At this stage the model is examined for consistency and interpretability. If an anomalous combination of variables has arisen then the pool of variables is adjusted to remove the anomaly and the process is then repeated. When a consistent model which replicates a previous model is finally derived then the model building process is complete.

Model 0 shows what sort of variables are important for explaining the construction processes; in particular, match structure and the number of properties known of the currently referenced individual. In order to model the data collected in this experiment, three new variables were created which reflected new aspects of the design. First, variables were incorporated to reflect the process of changing reference because it was expected that unpredictable shifts of attention would have an associated cost (in the variable naming convention SW means a switch of reference). Second, variables were distinguished by the status of the individual to which they referred because the recall data suggested that subjects were focussing one individual and therefore might be processing it differentially (PR means primary individual and SEC means secondary individual). Third, variables were added to reflect word length effects on reading time (SYL means number of syllables of a description). The development of these three groups of variables were not coincidental but were developed in stages as an understanding of the task increased. For example, it was not until the analysis of the recall data that status was incorporated and only then were the complex reading time effects revealed.

To capture the various aspects of referential change two new variables were defined. If reference had switched then FOREGROUND took as its value the number of properties which were known of the individual to which reference had switched but had been learned before the current sentence. If reference continued, the value of FOREGROUND was 0. FOREGROUND was differentiated by the status of the currently referenced individual so that there were two versions corresponding to the primary and secondary individual which were called PRFORE and SECFORE respectively.

MATLOAD and MISLOAD were also differentiated by the status of the currently referenced individual in order to further investigate the effects of status. However, NEUTLOAD which is directly related to the match structure of the individuals in a text was not differentiated by status. NEUTLOAD only takes on a non-zero value for a secondary individual in one sentence of one mode (Mode 5, Sentence 7) because the secondary individual, by definition, generally lags behind the primary individual. That is, on any dimension the primary individual's property is usually discovered first. Therefore, because there is only one non-zero position for the secondary individual there is not enough variation to reliably estimate any primary/secondary differences in the NEUTLOAD coefficients. As described above, these variables were added to the pool of variables as dummy variables.

Several variables were included in the general pool of variables to investigate word length effects. Basically two variables, REFSYL and NONREFSYL were defined which took on as their values the number of syllables in the currently referenced and non-referenced individual's description. For example, if a subject was reading the sentence "The square is green" and had already learnt that there was a square and that there was also a large blue parallelogram, then the values of REFSYL and NONREFSYL would be 2 (/green/square/) and 7 (/large/blue/pa/ra/lle/lo/gram/) respectively.

REFSYL and NONREFSYL were further differentiated by status, by switch of reference and by whether the syllables had accumulated prior to or since the current run of reference. For example, NONREFSECSWSYL took on the number of syllables in the non-referenced individual's description when there was a switch of reference to the primary individual (making the non-referenced individual the secondary individual), otherwise its value was 0. To illustrate the difference between runs of reference and switches of reference, REFRUNSYL took on the number of syllables which made up the description of the currently referenced individual which had been learnt since the last switch of reference. The syllabic variables took values between 0 and 13 which was the maximum number of syllables in any individual's complete description in the texts used. These variables were not recast as dummy variables because Baddeley, Thomson and Buchanan (1975) had found a linear relationship between word length and articulation rate.

A binary variable FORMAT was used to distinguish between the two formats, forward and backward. Because the ANOVA had shown that format only had a main effect on reading times and did not interact with any of the other variables a single binary variable was enough.

The model which was finally developed (and will be referred to as Model 1) is shown in Table 2.9 alongside Model 0 which was the label used for Stenning, Shepherd and Levy's (1988) model.

2.2.4 Discussion of regression model

The regression model developed for the data in this experiment is a close relative of the one developed by Stenning, Shepherd and Levy (1988) for data derived from a much simpler task. At the most basic level the current model is a replication of Model 0 with extra variables added to take account of the differences in the two tasks.

There are of course many features of the models which are worth noting but perhaps one of the most remarkable is the independence of the processes which is indicated by the high degree of fit of such a simple linear model. For example, in Model 0, the levels of MISLOAD each occur at several different positions in the two modes yet their coefficients provide a good fit regardless of the preceding processing history. In the current experimental design all levels of all factors occur at several different positions in the 7 modes and the definitions of the variables all assume that the effects of the variables are independent of the processing history. For example, MISLOAD may take the value 1 anywhere between the third and eighth sentence of a text and is constrained by its definition to have the same effect at all these positions.

Consider the observation that the NEUTLOAD function is similar in the two models which is impressive given the different circumstances. In the predictable texts (P$\times$P and I$\times$I), NEUTLOAD was only ever positive on sentences about the first introduced individual. However, in the current experiment, NEUTLOAD takes all of its values except the highest on both individuals. Once it is accepted the NEUTLOAD function is similar in both models, this replication demonstrates that the variable is robust and shows that the number of unresolved properties known of the referenced individual imposes a significant time cost whatever the processing history.

Status of Individual	Continues Reference	No. Properties of New Referent 0	1	2	3
Primary	0	0	0.18	1.22	1.65
Secondary	0	0	0.18	0.42	0.55

Table 2.11: Summary of effects of referential change in Model 1: total coefficients (centiseconds) for each property by continuity of reference and number of properties previously known of new referent and by primary/secondary individual

Neither MATLOAD nor LOCALMIS are included in the model for the current experiment (Model 1) which shows that they had no predictive power. Furthermore differentiating MISLOAD (which is included) by status did not improve the fit of the model which is in direct contrast to the importance of status for the process of changing reference as demonstrated by the PRFORE variable and the various word length variables. The MISLOAD variable is included in Model 1 but the shape of the function is different. In Model 0 the function was close to being linear with the number of mismatches in the referenced individual's description but in the current model the function is more like an n-shaped curve with a plateau when MISLOAD takes the value 2 or 3.

The absence of MATLOAD indicates that subjects relied more on the inherent redundancy in the individuals' descriptions than in Stenning *et al.*'s experiment. When a subject discovers a matched dimension then the subject only needs to remember that the dimension matched and which property it was that matched rather than integrating the properties into the representations for the two individuals. This is a simple short cut which subjects obviously discovered.

Although there are of course differences between Model 1 and Model 0 there is a high degree of fit considering the difference in design. The modularity of the processes and the importance of match structure are clearly replicated.

2.2.5 Switching reference in a text

In order to asses more easily what Model 1 shows about switching reference Table 2.11 presents the various summations of the relevant coefficients related to reference switching.

The table clearly shows that the more the reader knows about the individual being switched to the longer the sentence takes to read. The increase is not linear and the difference between one and two properties is about twice the size of the difference between two and three properties. Also, the slope is steeper for primary individuals than it is for secondary individuals: learning a new property of the primary individual takes more than three times as long as learning a new property of a comparable secondary individual. This shows that the process of switching reference is asymmetric and differentiated by status.

	Reading about Primary		Reading about Secondary	
	Referenced	Non-referenced	Referenced	Non-referenced
Reference continued	0.06/0.02*	0.04	0.22/0.18*	-0.13
Reference changed	0.06/0.02*	-0.26	-0.02/-0.06*	0.04

Table 2.12: Summary of syllabic effects in Model 1: coefficients (seconds) for each syllable of description accumulated since the beginning of the text, on referenced and non-referenced individuals, by continuity of reference and by primary/secondary individual. Note: In cells marked * there is a difference in the coefficients for syllables accumulating before the current reference was established (shown on the left), and those accumulating since (shown on the right)

2.2.6 Word length effects and the role of articulatory rehearsal

The regression model revealed various effects of word length on reading times. Table 2.12 summarizes these effects.

The table shows that reference switching and primary/secondary status interact to affect the word length effects for both referenced and non-referenced individuals. There are basically two types of effects represented by positive and negative coefficients. The positive coefficients reflect a positive relation between number of syllables and reading times showing that as the number of syllables increases so does the reading time. Negative coefficients reflect a negative relation indicating that as the number of syllables increases the contribution to the overall reading times decreases. The effects are differentiated by reference: there are effects which depend on the number of syllables accumulated in the current individual's description since the last switch of reference (called a *run of reference*) and there are effects which depend upon the number of syllables in the individual's description already discovered regardless of switches of reference.

Consider first the primary individual as the referenced individual (the leftmost columns in the table). Whether reference has been continued or whether it has switched has no effect on the (word length) effects for the referenced (primary) individual. The effects are fairly small and show that the description discovered during the current run of reference contributes a small effect (0.02 seconds) on top of the contribution made by the whole description (0.06). However, the non-referenced or secondary individual is affected by the history of referential change. If reference has been continued then the number of syllables in the secondary individuals description has a small and positive effect on the reading time (0.04 seconds). If reference has switched then the number of syllables in the secondary individual's description has the dramatic effect of accelerating reading by a fairly large amount (0.26 seconds faster).

Now consider the word length effects when the secondary individual is the currently referenced individual (the two rightmost columns in the table). The pattern is clearly quite different. When reference is continued the number of syllables in the secondary individual's description makes a large positive contribution (0.22 seconds) and the fraction accumulated since the switch of reference also makes a considerable contribution (0.18 seconds). However when the secondary individual has been switched to the number of syllables in its description only make a small negative contribution (-0.02 and -0.06 seconds). While the secondary individual is being read about the number of syllables in the primary individual's description is causing an acceleration (0.13 seconds) if reference has continued but if reference has switched then the number of syllables contributes a small positive time (0.04 seconds).

In summary, the number of syllables in the referenced individual's description usually adds to the reading time and the number of syllables in the non-referenced individual's description leads to a decrease in processing time when there is continuation of reference to the secondary individual or a switch of reference to the primary individual.

The positive syllabic word length effects are interpreted as articulatory rehearsal effects, as described in Chapter 1. The model of Working Memory described in Baddeley (1986) outlines a process, whereby visually presented words are encoded phonologically by the reader executing an articulatory program and are refreshed by repeatedly executing this program. These articulatory programs have the property that they are directly related to the time to say the corresponding words so that the longer a word is the longer it will take to refresh. It is on this basis that the increase in number of syllables causing an increase in reading time is interpreted as evidence for articulatory rehearsal and hence deployment of phonological representations in the type of memory system proposed by Baddeley (1986).

The negative word length effects which reflect an acceleration in processing are novel. They are related to articulatory rehearsal but what they represent is an acceleration of *other* process which is predicted by the amount of articulatory rehearsal needed to maintain the phonological representation of the description causing the acceleration. Consider a description which is being held in the phonological loop which is liable to decay and will be irretrievable if not refreshed. In between refreshing these items there will be a certain amount of time available for other processing. This spare time will depend on how long it takes to refresh all the items being stored because the items have a fixed decay constant: the longer it takes to rehearse all the items the less time there will be be to perform other processes. Therefore, these negative effects are interpreted as representing an acceleration of other processes in order to allow the phonological representation that is being maintained when they appear, to remain accessible.

There are, however, alternative explanations of these positive word length effects. Wright (1979) found that high frequency words were articulated more rapidly than low frequency words. Because the analysis of word length effects was not anticipated, frequency effects were not controlled for in the vocabulary and there is a negative correlation between word frequency and word length, $r = -0.40$, in the vocabulary (Colheart, 1981). The correlation is larger than the correlation for the whole MRC database (Coltheart, 1981) which is, $r = -0.18$. This means that on average, long words are less frequent than short words and therefore take longer to say which may explain the word length effects in this experiment.

In summary, this experiment has revealed a distinction between primary and secondary individuals and their attendant differences in processing related to switching reference. However, the range of modes used makes it difficult to discriminate certain effects and the assignment of primary/secondary status to the two individuals of a text is *post hoc*. The interpretation of the word length effects is also affected by the effects of frequency in the vocabulary. Therefore Experiment II was designed to solve these various difficulties and replicate Model 1.

	Sentence Position							
Mode	1	2	3	4	5	6	7	8
1	a 1	a 2	a 3	a 4	b 1	b 2	b 3	b 4
2	a 1	a 2	a 3	b 1	a 4	b 2	b 3	b 4
3	a 1	a 2	b 1	b 2	b 3	b 4	a 3	a 4
4	a 1	b 1	a 2	a 3	b 2	a 4	b 3	b 4
5	a 1	b 1	b 2	a 2	a 3	b 3	a 4	b 4
6	a 1	b 1	b 2	b 3	a 2	a 3	b 4	a 4
7	a 1	b 1	b 2	b 3	b 4	a 2	a 3	a 4
8	a 1	a 2	b 1	a 3	a 4	b 2	b 3	b 4
9	a 1	a 2	b 1	b 2	a 3	a 4	b 3	b 4
10	a 1	b 1	b 2	b 3	a 2	b 4	a 3	a 4

Table 2.13: Sentence sequences in each mode (letter denotes individual and number denotes property). Experiment I used Modes 1 to 7 and Experiment II used modes 1 to 10.

2.3 Experiment II

In Model 1 the combination of the FOREGROUND and PRFORE variables accounted for the long reading times when there was a switch of reference to the primary individual. For example, Table 2.11 shows that when the primary individual is switched to and three properties are already known then 1.65 seconds is added to the reading time.

However, the situation where a switch is made to the primary individual when only three properties are known only happens in two places: Sentence 5, Mode 2 and Sentence 6, Mode 4. In both situations there has been no change of status which means that the first introduced individual is the primary individual throughout the text. Inevitably the question arose as to whether the long reading times could be attributed to the individuals status or to its order of introduction. Given the importance and emphasis given to the asymmetric switching behaviour of primary and secondary individuals an extra mode was designed to contain a switch to an almost complete description of the primary individual after a change of status. This new mode acted as a test to see if a long reading time would still be found even though the primary individual had become the second introduced individual. This new test mode was Mode 10 and is shown in Table 2.13. Two other modes (Modes 8 and 9) were also added in which no change of status was expected so that the number of modes which did contain a change of status equalled the number of modes which did not.

An important aspect of Model 1 was the discovery of the word length effects. Because they had not been anticipated the correlation between number of syllables and number of properties within a description was higher than necessary and the correlation between frequency and word length was also high. Therefore, frequency needed to be controlled and the effects of word length and number of properties separated. The way to reduce the correlation between number of properties and word length is obviously to have a vocabulary which has a large range of word lengths. Unfortunately this would lead to unnatural descriptions of individuals because long words are inevitably more rare than shorter words. Therefore, frequency was controlled while decreasing the correlation between word length and number of properties as far as possible while maintaining natural descriptions.

2.3.1 Method

Subjects

Thirty one psychology undergraduates participated as part of a course requirement.

Materials and Design

Vocabulary

The vocabulary was required to be balanced for word length and frequency which is not possible with shape descriptions. Therefore the semantic field was changed to descriptions of people. Stenning, Shepherd and Levy (1988) had already demonstrated the descriptions of people and of shapes behaved similarly in the MIT. The dimensions used were similar to those of Stenning *et al.*: namely, Profession, Nationality, Stature and Temperament.

The vocabulary was split into two groups, of high and low frequency words. High frequency words were defined as occurring in 20 or more of the 500 samples of texts in the Francis and Kucera (1982) word count and low frequency words occurred in 10 or less samples. The two groups of high and low frequency words were further subdivided into groups of short and long words. Short words were one syllable long and long words were 2 or more syllables long. Unfortunately, there are not many single syllable nationalities in English (for morphological reasons; English usually requires an *-ish* or *-an* at the end of nationalities). *Welsh* was used as a high frequency item, although its rating in Francis and Kucera (1982) is low. It is almost certainly a high frequency item for Scottish speakers as opposed to American writers. Also, *Czech* was used as a low frequency item although it had no rating in Francis and Kucera (1982) because *Czechoslovakia* was given a rating of 4 samples out 500.

There were 48 words in the vocabulary arranged in contrastive pairs making 24 pairs. Twelve pairs were high frequency words and 12 pairs were low frequency words. Each group of 12 pairs was split into the 4 dimensions so that there were 3 pairs of words in each frequency group within each dimension. The 3 pairs were arranged so that one pair was made up of two long words (LONG—LONG), one of two short words (SHORT—SHORT) and one of one short and one long word (SHORT—LONG). The vocabulary is presented in Appendix A.1.2.

The texts were generated by a procedure similar to the one in Experiment I. All the vocabulary items in a text were either high or low frequency words. Apart from that, when the selection of pairs of items was made for each dimension, a pair of items was picked at random from the appropriate three pairs. This meant that frequency was controlled and that word length was randomised thus minimizing the correlation between word length and number of properties within a text.

Thirty one lists of materials of 50 texts each were constructed. Frequency, Mode and Matchtype were all design variables which applied to texts. Frequency and Mode were balanced and Matchtype was randomised. The order of texts within a list was randomised.

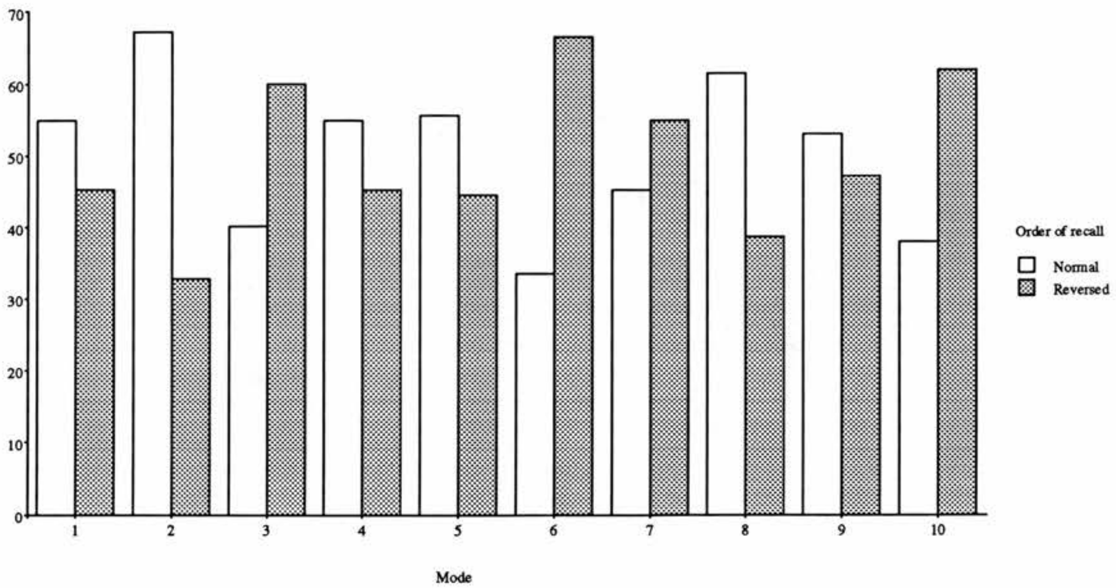

Figure 2.2: Percentage of subjects who recalled the individuals in the canonical and non-canonical order by Mode. In Modes 3, 6, 7, and 10 individuals were more frequently recalled in the inverse introduction order.

Procedure

The procedure was the same as Experiment I except that all the texts were read within one sitting. During the recall phase of the experiment subjects were not cued as to the order in which the individuals were to be recalled in order to keep the size of the design reasonable.

2.3.2 Results

Recall

The recall was scored by the order in which the individuals were recalled. Figure 2.2 shows the percentage of texts which were recalled in the order in which the individuals were introduced and the reverse order.

The table clearly shows a pattern of results which fits very well with the two groups of modes, differentiated by whether a change of status takes place. In Modes 3, 6, 7 and 10 the two individuals are more often recalled in the reverse order of introduction and in Modes 1, 2, 4, 5, 8 and 9 the two individuals are more often recalled in the introduction order. This is exactly the pattern of results which would be expected if subjects were recalling the primary individual first. The subjects were not instructed to recall the individuals in any particular order so the fact that they preferred to recall the primary individual first supports the notion that the individuals were ordered by status in some way. Perhaps subjects preferred to recall the individual which they thought had been the first introduced individual which they assumed to be the primary individual or perhaps because they were "focussing" on the primary individual more than the secondary individual, "it came to mind" first. Of course, these speculations cannot be confirmed with the current data set but the pattern of results at least conforms

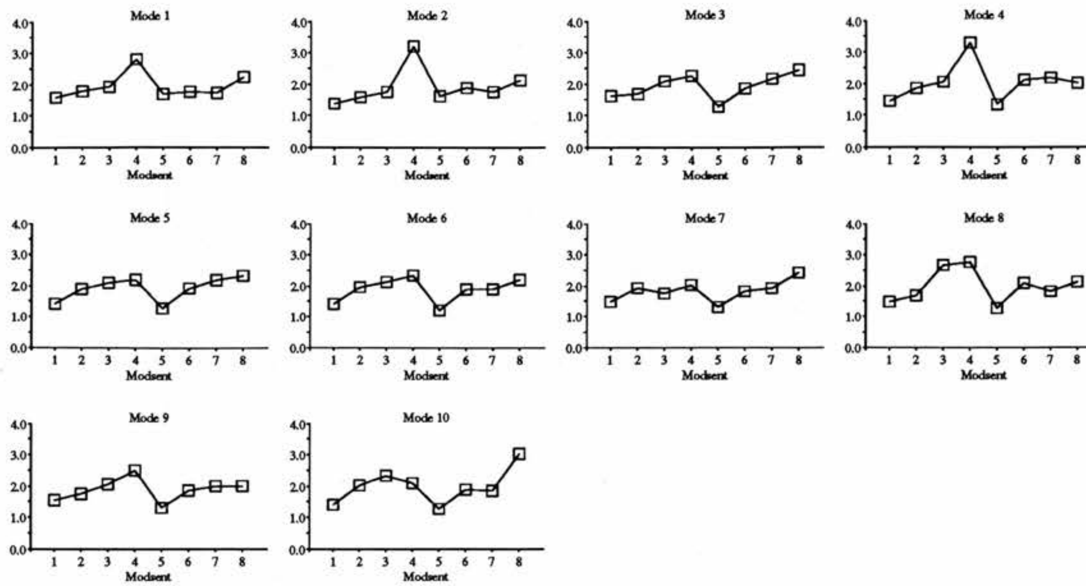

Figure 2.3: Experiment II mean reading times (sec) as a function of Modsent and Mode.

with the pattern found in Experiment I. Furthermore Mode 10 behaves in the same was as the other Modegroup 2 modes and Modes 8 and 9 behave like Modegroup 1 modes, which lends support to the rule for assigning status which was used to design them.

The exception to the pattern is Mode 5 which, according to the rules, does have a change of status, and yet behaves like a Modegroup 1 mode. The explanation may lie in the possibility that there are two changes of status in Mode 5 and that by the time a text has been read the primary individual is once again the first introduced individual.

Reading Times

Figure 2.3 shows the mean reading times by Modsent and Mode.

Regression modelling

The data was investigated using the same procedure as in the analysis of Experiment I. The definitions of primary and secondary individuals was the same for Modes 1 to 7 except for a change which affected Mode 3. In Experiment I, there was a change of status at Sentence 4 in Mode 3 but in this experiment the change of status took place at Sentence 5. This change was caused by a different pattern in the reading times so that the rule which had been added to take account of the change in status at Sentence 4 of Mode 3 was replaced by a rule which determined that the status of the individuals changed if B3 preceded A3. Consequently, the new rule predicted a change of status at Sentence 5 of Mode 3.

The pool of variables offered to the stepwise regression procedure was the same as the pool used in the previous regression modelling analysis with the following exceptions. Of course, FORMAT was not applicable in this experiment. CONTOUR had been in the pool of variables used in the analysis of Experiment I. It is a binary variable which takes the value 1 when there

Model	Model 2		Model 1	
	$R^2 = 0.121$	DF. = 20	$R^2 = 0.123$	DF. = 22
Variable	Coeff(sec)	St. Err.	Coeff(sec)	St. Err.
INTERCEPT	1.58		1.00	
FREQUENCY	0.09	0.02	n/a	n/a
FORMAT	n/a	n/a	0.25	0.03
NEUT1	—	—	0.29	0.06
NEUT2	0.21	0.04	0.30	0.08
NEUT3	0.47	0.06	0.45	0.10
NEUT4	1.32	0.11	1.18	0.15
MIS1	0.26	0.05	0.33	0.06
MIS2	0.60	0.06	0.53	0.07
MIS3	0.76	0.07	0.53	0.09
MIS4	0.69	0.10	0.42	0.14
CONTOUR	-0.58	0.07	—	—
FORE1	0.15	0.04	0.18	0.07
FORE2	—	—	0.42	0.11
FORE3	—	—	0.55	0.15
PRFORE1	0.83	0.12	—	—
PRFORE2	1.47	0.11	0.80	0.18
PRFORE3	1.90	0.13	1.10	0.17
NON-REFPRSWPROP	0.17	0.02	—	—
NON-REFSECSWPROP	-0.21	0.06	—	—
REFSYL	0.05	0.01	—	—
NON-REFSYL	0.07	0.02	-0.18	0.04
NON-REFPRSYL	-0.09	0.02	-0.17	0.03
NON-REFSECSWSYL	-0.15	0.04	-0.47	0.04
REFRUNSYL	-0.04	0.01	-0.04	0.02
REFPRSYL	—	—	-0.16	0.02
REF+NON-REFSYLL	—	—	0.22	0.03
REFSECSWSYL	—	—	-0.24	0.03
NON-REFSWSYL	—	—	0.17	0.02

Table 2.14: Summary of reading time Model 2 with Model 1 repeated for comparison

is a switch of reference and 0 on all other occasions.

Table 2.14 shows the model which was finally derived and will be referred to as Model 2. The following convention for naming variables is used: REF=referenced, NON-REF=non-referenced, PR=primary, SEC=secondary, SW=switched reference, RUN=accumulation since reference switch, SYL=syllables, PROP=properties. For example, NON-REFSECSWSYL refers to the number of syllables in the secondary individual's description when it is the non-referenced individual after a switch of reference.

Model 2 is very similar to the previous model (Model 1) although there are clear differences in the word length effects. The NEUTLOAD function still has a very similar shape in Model 2 although the dummy variable representing NEUTLOAD's coefficient at 1 is absent. Presumably this difference must be related to the new modes where the occurrences of a NEUTLOAD of 1 (11) contribute a third to the total instances (34). They occur at positions where other variables contribute a large reading time so NEUT1 positions may have become too noisy to distinguish. Of the three levels of NEUTLOAD which remain the size of the coefficients and the shape of the function are very similar in the two models. The function for mismatched

Status of Individual	Continues Reference	No. Properties of New Referent 0	1	2	3
Primary	0	-0.58	0.99	1.47	1.90
Secondary	0	-0.58	0.15	0	0

Table 2.15: Summary of effects of referential change in Model 2: total coefficients (seconds) for each property by continuity of reference and number of properties previously known of new referent and by primary/secondary individual

properties (MISLOAD) shows a similar shape in Model 2 to that in Model 1 although the curl-over is less marked and the plateau in observed in Model 1 is absent. As in Experiment I the number of mismatched properties has no effect on the reading time.

Switching reference

The process of switching reference is similarly represented in Model 2 as in Model 1. Table 2.15 presents the analogous table to Table 2.11 which presents the coefficient summations relevant to switching reference.

The table shows that switching to a new individual actually accelerates reading compared to continuing reference to an old individual (0.58 seconds). There is also a clear difference between primary and secondary individuals in the way they affect reading time at a switch of reference. For switches to primary individuals the more that is known about the individual the larger the contribution to the reading time (-0.58 to 1.90 seconds): this pattern replicates the effect found in Experiment I. The secondary individual is quite different. If one property is known then the reading times is increased slightly (0.15 seconds) but as more is known switching to the secondary individual costs no time. The process of switching reference has slightly changed between the two models although the main characteristics are replicated: switching to an established primary individual costs time and the process of switching reference is different for primary and secondary individuals. Given that in this experiment the modes were more evenly balanced and that there were more subjects than in Experiment I it is likely that switching reference only costs time when switching to a primary individual.

Recall that Mode 10 was introduced to test whether long reading times would results when switching to a primary individual which was the seconds introduced individual. In Mode 2, Sentence 5 and in Mode 4, Sentence 6 there is a switch to a primary individual about which 3 properties are already known and those sentences replicate their long reading times (3.20 and 3.32 seconds respectively) found in Experiment I. In Mode 10, Sentence 6 there is a similar switch to a primary individual which is the second individual and it produces the predicted long reading time (3.02 seconds). This confirmation of a prediction provides justification for the distinction between primary and secondary individuals introduced in the recall analysis of Experiment I. The distinction is not simply correlated with order of introduction and the concept which was derived from recall analysis has predictive power for reading time analysis.

	Reading about Primary		Reading about Secondary	
	Referenced	Non-referenced	Referenced	Non-referenced
Reference continued	0.05/0.01*	0.06	0.05/0.01*	-0.03
Reference changed	0.05/0.01*	-0.09	0.05/0.01*	-0.03

Table 2.16: Summary of syllabic effects in Model 2: coefficients (seconds) for each syllable of description accumulated since the beginning of the text, on referenced and non-referenced individuals, by continuity of reference and by primary/secondary individual Note: In cells marked * there is a difference in the coefficients for syllables accumulating before the current reference was established (shown on the left), and those accumulating since (shown on the right)

Word length effects

Table 2.16 summarizes the word length effects in Model 2. These effects are substantially different to the effects in Model 1 although there are important similarities. The largest effect is still the acceleration in reading time in proportion to the number of syllables in the secondary individual's description when there is a switch of reference to the primary individual. In general the pattern of effects when reading about the primary individual have remained the same. When reference is continued to the primary individual, the number of syllables in both individuals' descriptions contributes a positive time and the number of syllables in the primary individual's description accumulated since the last switch of reference contributes a small but significant amount. When reference has switched to the primary individual the word length effects of the primary individual's description are the same as those when reference has continued. However, as described above, the secondary individual's description accelerates reading.

The most substantial changes between the two models are found in the word length effects when reading about the secondary individual. The effects for both individuals are the same regardless of referential continuity and are the same for the referenced individual when reading about the primary individual (positive effects for the whole description as well as a small contribution from the description accrued since the last switch of reference). There are small negative effects of the primary individual's description which are different from the effects in Model 1 where there was a small positive effect when there was a switch of reference and a large negative effect when reference continued.

The most surprising effect in Model 1 is the negative effect of the secondary individual when there is a switch of reference to the primary individual. Frequency is obviously not accounting for it because the effect remained in Model 2 when FREQ had been added. However, the correlation between number of properties and number of syllables in a description is quite high, so a new variable (NON-REFSECSWPROP) was added to represent the number of properties in the secondary individual's description when there is a switch of reference to the primary individual to see if it could account for the negative word length effect. Model 2 shows that it does not, although it does enter the model with a negative coefficient.

2.4 General Discussion

It is clear that Model 2 replicates much of the description of the reading times processes presented in Model 1. The description is split broadly into two groups of processes: semantic and rehearsal (as indicated by word length effects). The semantic processes represented by NEUTLOAD and MISLOAD and the absence of MATLOAD are replicated in Model 2. The notion of status and its relation to switching reference is also replicated in the semantic processes by the presence of PRFORE and CONTOUR and in the rehearsal processes by the negative word length effect when there is a switch of reference to the primary individual.

However, there is a substantial change in the word length effects in spite of the replication of the negative effect mentioned above. There could of course be little rehearsal taking place but then it would be less likely that the negative effect would be replicated. An alternative explanation lies in the possibility that more rehearsal was taking place in parallel with the semantic processes so that the rehearsal of descriptions was not a delaying factor in reading and therefore did not enter the regression model with the same coefficients as in Model 1. This is consistent with the fact that the property load variables and the intercept in Model 2 are higher than in Model 1 which suggests a greater emphasis on semantic processing.

In summary then, these two experiments have shown that Stenning, Shepherd and Levy's model of reading time processes applies to unpredictable modes in that similar semantic processing takes place. However, unpredictable modes cause the two individuals to be treated differently: one is "focussed" on more than the other and the notion of a primary and secondary individual was introduced as a result of identity confusions in the recall phase of the task. This notion of status is particularly important for the semantic processing when switching reference because it shows that switching reference is an asymmetric process. The new notion of status also revealed word length effects, which were interpreted as articulatory rehearsal phenomena, and they too are closely linked with status and switching reference.

There are two issues remaining to be discussed: why have primary and secondary individuals emerged in unpredictable texts and what role is articulatory rehearsal playing in the construction of the representation of individuals? A discussion of the role of articulatory rehearsal will be postponed until Chapter 4 where its use in the MIT will be discussed followed by its investigation.

2.4.1 The purpose of status

Recall that the sentence-by-sentence rule used to assign primary/secondary status was:

1. Assume that the first introduced individual has primary status and that the secondary individual has secondary status.
2. If B2 (the second property of the second individual) precedes A2 (the second property of the second introduced individual), then change the status of the individuals.
3. If B3 precedes A3 then change the status of the individuals.

and that it divided the 10 modes into two groups, those which contained a change of status (Modegroup 1) and those which did not (Modegroup 2). The division was determined by the recall and reading time data and had been replicated and tested. Further, the distinction was not needed for the predictable modes investigated by Stenning, Shepherd and Levy (1988). Therefore the unpredictably introduced in the new modes must have caused the introduction of status. The unpredictably introduced applied to the sequence of reference but this also meant that for any dimension (other than the introducer) the order in which the individuals became described on that dimension became unpredictable.

In Modegroup 1, for any dimension the first mentioned individual is the first introduced individual. For example, if the dentist is introduced first followed by the bishop in Mode 4 then the Stature of the dentist will be mentioned (at Sentence 4) before the Stature of the bishop (Sentence 7). However, in Modegroup 2 this is not the case. For example, in Modes 6, 7 and 8 the order in which the individuals are mentioned on Dimensions 2-4 (Nationality, Stature and Temperament, in Experiment II) is the reverse of the introduction order. Suppose, for example, that the dentist is the first introduced individual and the bishop the second introduced individual then in Mode 6 the Stature of the bishop will be learnt (Sentence 4) before the stature of the dentist (Sentence 6). Consider now the order of mention within a dimension by the status of the individuals. In all modes, except Mode 5, the order of mention is constant: primary individual first then secondary individual. It is apparent that at least one of the effects of the primary/secondary distinction is to restore constancy to an aspect of the organization of the experimental texts. Mode 5 is the exception but this is not surprising because the diagnostics of a status change (order of introduction confusions and reading times) are unreliable for it. It would be reasonable to conclude that there are perhaps two changes of status in Mode 5 which would account for the recall errors and allow the primary/secondary distinction to ensure that the first mentioned individual on any dimension was the primary individual.

Of course the obvious question is still how does the primary/secondary distinction help the reader? In predictable texts there is a simple mapping from surface structure to semantics. All the reader has to do in the MIT is remember which property on a mismatched dimension came first and which came second and that is enough to remember which property went with which individual. In the unpredictable texts used in these two experiments, the mapping has been destroyed, but it can be restored by using primary and secondary individuals. Then, all the subjects have to do is remember that the first property on a dimension applies to the primary individual and that the second property applies to the secondary individual. Therefore the purpose of the primary/secondary distinction is to restore the transparent mapping between within-dimension order and semantics.

2.4.2 Implications for memory

The primary/secondary distinction manifests itself in the concurrent processing of a text and in its (the text's) subsequent recall. Clearly, working memory is involved in this task (the MIT) and the question remains: what is the relationship between the primary/secondary distinction and the representation of individuals in memory?

First of all, is there any evidence that within-dimension order is encoded by subjects? Stenning (1991) presents a reanalysis of data which demonstrates that there is such evidence. This evidence depends upon the way that the vocabulary is constructed for use in MITs. As in the experiments described in this chapter, the vocabulary is composed of pairs of properties, one of which is often marked and the other which is often unmarked (see Clark, 1969, for a description of marking). Usually it is the case that the pair of properties have a natural *citation* order which is the unmarked property followed by the marked property, for example, *fat/thin*. This citation order is reinforced by the use of settings which precede texts and if there is no inherent citation order then the setting serves to introduce a sense of ordering over pairs of properties. This sense of ordering is used as a tool to investigate effects of within-dimension order on recall and on reading times. It is assumed that it is easier to remember a pair of words if they appear in canonical (Citation or Setting) order as opposed to non-canonical order. If the representation of within-dimension order is important then it would be expected that effects of citation order would appear. The actual predictions are fairly complicated and interact with match structure because citation order has no meaning for matched dimensions.

However, Stenning (1991) found that when the mapping between surface order and semantics had been disrupted (as in a Modegroup 2 modes), counter citation-order of a dimension slowed processing. Furthermore, similar effects of citation order were found in the recall results although the difference between texts from the two Modegroups was not significant, as in the reading time analysis. Although this analysis was a *post hoc* one it shows that within-dimension order has interpretable effects which do support the role of the mapping between superficial properties of a text and its semantics.

Stenning, Shepherd and Levy (1988) and Stenning and Levy (1988) describe a representational scheme for solving the binding problem, as discussed in Chapter 1. Essentially the solution lies in making an inference from a set of informationally-redundant existential facts about the individuals to a complete pattern of binding. In the parallel distributed processing (PDP) model described in Stenning and Levy (1988) there are fifteen facts, which summarise local aspects of the pair of individuals, for example, "There is an individual who is fat and there is an individual who is not fat". These facts are represented by input nodes in a feedforward 3-layer network. The output of the network is the complete pattern of binding in the form of eight nodes which can represent both individuals, one node for each dimension for each individual. The network simulates recall of a pair of individuals by clamping an appropriate pattern of input nodes, allowing activation to feed forward and reading off the networks inference on the output nodes.

This view of the recall process and representational system sees the low level facts as being a trace which is inaccessible to conscious recall. The recall process is an active inference from the set of facts. One major advantage of such an arrangement is that the system can always produce an answer even when the representation has been corrupted and a logical inconsistency has been produced. The appearance of primary/secondary status in unpredictable texts shows that it is important for the processor to restore the mapping between surface intra-dimension order and semantics. The reason why it is important could be because the low-level representational system encodes information about sequence which is essential for the inference process to take place. It is possible that the mechanism for inferring accessible information from lower-level inaccessible information is a general one which needs to have its input sorted in a particular fashion and in this instance the primary/secondary distinction achieves this sorting.

The discovery or revelation of the primary/secondary distinction has interesting implications for short-term memory. What it implies is that in this task the memory system used assumes a particular sort of mapping between surface order and semantics. When texts are made unpredictable and this mapping is disrupted the processor restores the mapping by introducing primary and secondary individuals. It may be the case that this inference mechanism, which depends in part upon a representation of order information, is in fact a general mechanism. If this is the case then it would be expected that in some other text processing situations evidence could be found which could be explained in terms of primitive encoding of serial order and it is this idea which will be investigated in Chapter 3.

Chapter 3

Parallelism in pronoun comprehension

Summary

The study of unpredictable reference change in simple texts revealed the use of parallelism which appeared to be a representational primitive. If this is the case then one might expect it to be observed in other aspects of text processing. In the pronoun resolution literature parallelism has been investigated to some extent although there have been serious confounds in much of the work. When these confounds are removed then parallelism does emerge as a strategy used by people in selecting an antecedent for a pronoun. This effect lends support to the proposition that parallelism is a general representational primitive.

3.1 Introduction

THE RESULTS from the experiments reported in Chapter 2 coupled with Stenning's (1991) analysis showed effects of ordering in text processing. Specifically the claim was made that order information was retained in a "low-level" representation which could be used in the general inference process used to recover conscious knowledge of patterns of binding. The imposition of status (the primary/secondary distinction) was interpreted as a way of restoring parallelism to the texts because first mentioned properties belonged to the primary individual and second mentioned properties belonged to the secondary individual just as first mentioned properties had belonged to the first introduced individual and second mentioned properties had belonged to the second introduced individual in predictable texts. These mappings are parallel in that the mapping from within dimension order onto individual order is parallel or identical across dimensions. The most direct evidence for this interpretation came from Stenning's analysis using marking which revealed order of mention effects in reading time data and in recall data. Stenning proposes that this low level information is encoded in representations as a matter of course and should therefore be available in other text processing tasks. Therefore it would be expected that order of mention effects might emerge in other tasks.

Gernsbacher and Hargreaves (1988) studied an effect of ordering which they called the "advantage of first mention." In their experimental task subjects read sentences which described a pair of individuals referred to by first names. After the sentences had been read subjects were required to verify a probe word (one of the first names) and their verification latency was faster for individuals mentioned first. They were concerned to show that the effect was independent of agency because agents are often mentioned first in English text (Greenberg, 1963). Therefore they constructed a set of materials in which agents were mentioned first half the time and second in the remaining materials. By using probes words for the first and second mentioned participants they demonstrated that the advantage of first mention remained irrespective of the agency of the first mentioned participant.

They also showed that the effect did not depend on a name being the first item of a sentence by manipulating the position of an adverbial phrase in an active sentence. When the adverbial phrase was preposed the named participants were no longer the first words of the sentence and the effect of first mention remained. Furthermore, they showed subjecthood was not responsible for the effect (subjects are often mentioned first in English sentences). They manipulated the subject status of the participants by making them conjoined subjects and still showed an effect of first mention.

Gernsbacher, Hargreaves and Beeman (1989) attempt to resolve the apparent conflict between advantage of first mention phenomena (described above) and recency effects in text comprehension. Recency effects are manifest in experiments where subjects read or hear a two clause sentence and find that words in the most recently processed clause or more accessible then words from the earlier clause. They propose a resolution using, what they call, a structure building framework. Each clause in a multi-clause sentence is assumed to have its own substructure. The items in a substructure that is currently being developed are the easiest to access which accounts for recency effects. However, the first clause does become more accessible because its representation or substructure is used as a foundation for the sentence-level representation. This increased accessibility accounts for primacy effects.

The experiments reported in Gernsbacher *et al.* (1989) attempt to test these proposals. They used two-clause sentences with two named participants whose accessibility they measured using verification latencies to probe words. They manipulated the order of mention of the probe word and time elapsed before the probe word was presented. Their findings showed that when the probe word coincided with the last word of the stimulus material there was a recency effect: probes which had appeared in the second clause were responded to faster than probes which had appeared in the first clause. However, 2000 ms after the sentence had been presented there was a primacy effect. They also investigated latencies 150 ms after the sentence (no difference) and at 1400 ms (primacy effect). Further experiments showed that if the participants were mentioned within the same clause then there was an advantage of first mention in spite of any clausal recency. They interpreted these results as supporting their structure building framework and were able to explain the primacy effect within a clause by simply proposing that the first mentioned individual served as a foundation for the clauses substructure.

Gernsbacher *et al.* (1989) see their structure building framework as a general framework not specific to linguistic stimuli. It so happens that nouns are a good foundation for substructures and clauses provide good units for substructures. The reason these studies are relevant here is because they demonstrate similar order phenomena which are not accounted for by attributes like agency or subjecthood but by representational properties which are assumed to be general. If representations commonly encode primacy and to some extent recency information then order information is available "for free" to linguistic processes and might therefore be expected to be used in normal language processing.

Work in the anaphora resolution literature on parallel function has identified the use of order information as well as function information in selecting an antecedent for an anaphor (Cowan, 1980). The phenomenon of parallel function is also interesting because it is an example of a simple heuristic applied to comparatively low level information: if something has the same function then it is likely to co-refer. If such parallelism exists then it might be expected that parallelism may operate over other primitive information like order-of-mention. Furthermore, grammatical parallelism is related to parallelism over order-of-mention because subjects usually come before objects in English.

The notion of parallel function was first proposed by Sheldon (1974) in a study on the acquisition of relative clauses. She found that if two identical nouns in a sentence with a relative clause (one in the main clause and one in the subordinate clause) had identical grammatical functions then the sentence was easier to understand than if the functions differed. For example, 3.1(a) is easier than 3.1(b).

(3.1) a. The dog that jumps over the pig bumps into the lion.

b. The lion that the horse bumps into jumps over the giraffe.

She speculated that parallel function was used in pronoun resolution, contrasting the following two examples (3.2a and 3.2b). In both cases the reference of the pronoun is determined by gender but in 3.1(a) the sentence is easier or more natural because the pronoun and its antecedent have the same grammatical function. In 3.1(b) the pronoun and antecedent have different grammatical functions and so the sentence reads less well.

(3.2) a. Mary hugged John and Betty kissed him.

b. Mary hugged John and he kissed Betty.

A series of experiments have been aimed at disentangling the effects of parallel function, implicit causality and topicalisation in pronoun assignment (Caramazza & Gupta, 1979; Garvey & Caramazza, 1974; Grober, Beardsley & Caramazza, 1978). Implicit causality referred to the property of some verbs which seemed to have a "direction of causality" that attributed the cause of an event to either the subject or object of the clause in which the verb appeared. The verb in example 3.3(a) carries a bias towards the object and the verb in 3.3(b) carries a bias to the subject. However, there is some doubt whether it is really the verb in isolation which determines bias. In example 3.4 the bias for the verbs seems to be reversed compared to the biases in example 3.3 although the verbs in the two examples are the same. However, Garvey, Caramazza and Yates (1976) explained the bias in terms of presuppositions which the verbs carried as part of their meanings. The causal bias of a verb determines the choice of antecedent in a completion task (Garvey *et al.*, 1975) and also facilitates the processing time in a timed comprehension task (Caramazza, Grober, Garvey & Yates, 1977).

(3.3) a. George_1 telephoned Walter_2 because he_1 wanted some information.

b. George_1 criticised Walter_2 because he_2 misplaced the file.

(3.4) a. George_1 telephoned Walter_2 because he_2 had asked for some information.

b. George_1 criticised Walter_2 despite his_1 usually placid temperament.

Grober *et al.* (1978) investigated parallel function, in the sense of Sheldon (1974) and proposed that parallel function was used in pronoun assignment when the available semantic cues did not provide an unambiguous antecedent.

Caramazza and Gupta (1979) provide the latest account of a particular attempt to investigate the three factors: parallel function, implicit causality and topicalisation. They followed Halliday (1970) in assuming that the first content word of sentence was the theme or topic of the sentence. They were also interested in the effects of backward and forward pronominalisation because they thought that there might be a difference in the effects of implicit causality: if the subordinate clause with the pronoun was preposed then they assumed that an assignment would be made using parallel function by the time the main clause was encountered and that implicit causality would have no effect. As part of their attempt to investigate parallel function they manipulated the voice of their materials because a sentence in the two voices would have the same semantics but opposite grammatical functions.

In their first experiment they used 10 verbs in four conditions (two experimental and two control) to make 40 sentences containing a main verb with two names and a subordinate clause with a subject pronoun. The experimental sentences contained two names of the same gender, making the pronoun ambiguous, and the control sentences contained two names of different gender, making the pronoun unambiguous. The subordinate clauses in the two types of sentence (experimental and control) fell into two conditions. Either they contained information that was consistent with the main verb's bias 3.5(a) or information that was inconsistent 3.5(b). Again the difficulty of attributing bias to the main verb in isolation can be seen in these examples. If "neat" is replaced by "horrible" in example 3.5(b) then the bias is reversed and the sentence becomes consistent.

(3.5) a. Because she_2 always looked so neat Ann_1 envied Mary_2.

b. Because she_1 never looked as neat Fay_1 envied Vicki_2.

Subjects saw all 40 sentences in a random order and were required to read each sentence and to respond by calling out the name which, in their judgement, was pronomilised. Reaction times and responses were recorded. Subjects were quicker for control sentences than for experimental ones and were also faster for consistent sentences than for inconsistent ones. Subjects preferred to make subject assignments across all conditions. However, there was some variation across the 10 verbs. Therefore they concluded that they had found weak evidence for parallel function because subject pronouns had been assigned to subject nouns and that their manipulation of preposing the subordinate clause had failed to remove the effects of implicit causality.

Their second experiment used the same materials and task with a slight modification. The main clauses were transformed to the passive voice and put in the initial position. They predicted that a parallel function strategy and a topicalisation strategy would agree on a subject assignment strategy which would override any bias in the main verbs in experimental sentences (ambiguous pronouns). They found a subject assignment strategy which confirmed their predictions. They also noted that their data could not distinguish between a topicalisation strategy and a parallel function strategy. Their third experiment was designed to separate these two strategies.

They compared the results from their first two experiments and deduced that passivisation caused an increase in subject assignments, in spite of verb biases which were attributed to effects of topicalisation. They further reasoned that if the pronoun was topicalised by using a preposed subordinate clause before a passive main clause there should be more subject assignments compared to an initial mention main clause if topicalisation has more influence than parallel function. However, if parallel function is the dominant strategy then there should be no effect on assignment bias of subordinate clause position in relation to a passivised main clause. Therefore they carried out a third experiment using the same task and materials with the alteration that the main clauses were passivised and the subordinate clauses preposed. The results again showed a subject assignment bias of the same magnitude as in their second experiment with passive main-clause-first materials. Unfortunately this did not lead to support of the parallel function hypothesis because the action of some of the verbs which had previously shown a bias to object assignment changed and caused a drift to a subject assignment bias.

In summary, the main result of their investigation was that subjects preferred to assign subject pronouns to subject antecedents. They observe that grammatical subjects are often confounded with other roles such as topic but fail to mention simple order-of-mention effects. So apart from confounding subjecthood with other semantic roles like topic, contrastiveness, givenness, *etc.* (Chafe, 1976) they also ignore simple surface ordering effects. Therefore their main result can only be interpreted as providing evidence that grammatical subjecthood, topicality and first mention may all bias assignment in first mentioned, topicalised subject pronouns. They do recognise this weakness and propose a solution to these problems by suggesting the study of non-thematic object pronouns.

Cowan (1980) noted that Caramazza, Grober, Garvey and Yates (1977) had investigated the effects of causal bias on anaphora resolution and pointed out that the characterisation of the verbs used was inaccurate. Furthermore he supposed that resolution strategies for intrasentential anaphora would operate independently of lexical features citing the work of Carroll and Bever (1976) and Clark and Clark (1977) among others. He proposed that parallel function was a more likely candidate as a strategy for resolving ambiguous anaphora and therefore investigated the effects of various structural properties of sentences on the use of parallel function.

He investigated four factors which might affect the use of parallel function. They were, order of mention of the potential antecedents, movement transformations, connectives between clauses and the influence of pragmatics. Example, 3.6 shows examples of the order of mention manipulation where the order of *measuring stick* and *stake* are contrasted.

(3.6) a. Maria positioned the measuring stick directly opposite the stake and then she lined it up with the reflector.

b. Directly opposite the stake Maria placed the measuring stick and then she lined it up with the reflector.

Because parallel function involves the identification of grammatical functions and transformations are known to disrupt the superficial clues necessary for analysis, three transformations were used: direct object passivisation, indirect object passivisation, dative movement and adverb preposing. Furthermore, all the nouns in most of the sentences were transformed using Chambers' (1979) letter distortion technique to render nonsense words so that subjects' knowledge of individual lexical items did not influence their choices. He investigated the effect of different connectives like *as* by including sentences like (3.7).

(3.7) Tom ran into Bill as he was rounding the corner in the hall.

and sentences with conjoined antecedents with different connectives (example 3.8) which the parallel function hypothesis should not be able to distinguish because the grammatical function is the same for both potential antecedents.

(3.8) a. Tom and Jim walked into the room, and he took off his hat.

b. John and Fred only played a few hands of poker that evening because he was tired and wanted to go to bed early.

The final factor of pragmatics was manipulated using sentences like (3.9). Clearly a rug could be painted but it would be unusual. Sentences like 3.9(b) were included to make sure that position was counterbalanced.

(3.9) a. Charles pulled the rug over to the bookcase before he began to paint it.

b. Charles pulled the bookcase over to the rug before he began to paint it.

The task required subjects to read booklets of experimental sentences and ring their choice of antecedent for the pronouns. The results showed a clear effect of parallel function over surface structure (order of mention) and a clear effect of pragmatics over parallel function. The transformations had variable effects. Direct object passivisation and adverb preposing did not disrupt parallel function but indirect object passivisation biased selection to the indirect object (parallel function would predict the direct object) and dative movement reduced any bias to either potential antecedent. The connectives did show that some disruption of parallel function was possible in the cases where coreference had to be established between a main and subordinate clause whereas connectives like *but* and *and* enhanced parallel function. In the cases where the two antecedents had the same grammatical function choices were made on a random basis or subjects indicated that the referent was a third person not mentioned in the stimulus sentence. Finally, Cowan noted that there was a bias across all his results to assign the pronoun to the first mentioned antecedent which although weak was noticeable. He concluded that parallel function is a psychologically valid strategy which can be overridden by various factors manipulated in his experiment.

Frazier, Taft, Roeper, Clifton and Ehrlich (1984) investigated the effect of parallel structure in sentence processing. They noted that in sentences like (3.10), the preferred parses of the sentence produced parallel structures for the two conjuncts rather than the interpretation which conjoined the NP, *the girl with a book* and the NP, *the boy*.

(3.10) Joshua hit the girl with a book and the boy with a bat.

Given that parallelism seemed to them a prevalent aspect of sentence construction they wished to investigate what levels of representation contributed to parallelism and how structural parallelism affected sentence processing. Therefore, they presented conjoined sentences and recorded subjects' reaction times for the two conjuncts while manipulating the structure of the segments.

Frazier *et al.* manipulated three factors which they called Sentence construction, Parallelism and Markedness. They contrasted five different Sentence constructions, active *vs.* passive, minimal attachment *vs.* nonminimal attachment, heavy NP shifted *vs.* non shifted, agent *vs.* theme and animate *vs.* inanimate. The first two contrasts (voice and minimal attachment) were designed to see if effects of parallelism only operated at points of temporary ambiguity (the voice sentences were unambiguous and the minimal attachment ones were ambiguous between direct object and sentential complements). The shift contrast was designed to determine at what level of syntactic representation parallelism operated, because according to Chomsky and Lasnik (1977) shifted and unshifted constructions' representations differ at the surface level but are identical at the S-structure level. The last two contrasts investigated any possible effects of thematic role. If parallelism only operated over constituent structure then thematic role should have no effect. The animacy contrast was included to avoid the confounding effect of animacy in the agent *vs.* theme contrast. The parallelism factor simply determined whether the constructions in the two conjuncts were the same or not and the Markedness factor determined whether the construction in the second conjunct was marked or not. For the purposes of the experiment, the unmarked construction were designated as, active, minimal attachment, unshifted, agent and animate.

The reaction times for the second segment of each experimental sentence was used in their analysis. They found that parallel sentences were read faster than non-parallel ones over all constructions. The higher order interactions were not significant on subjects and items analyses but there were indications of possible effects. However they conclude that the data does not distinguish between the two possible interpretations of parallelism which they advance: namely, that parallelism increases the availability of structural information in general or whether it arises from a variety of distinct mechanisms. In spite of their disappointing conclusions their results do show the structural parallelism has a measurable effect across a variety of constructions.

Crawley, Stevenson and Kleinman (1990) investigated the use of parallel function and a subject assignment heuristic in pronoun interpretation. They noted that previous work which had attempted to separate these two strategies had been flawed because only subject pronouns had been used which could not reliably separate these two strategies. They also noted three studies which had used object pronouns (Maratsos, 1973; Fredriksen, 1981; Rondal, Brédart, Leyen, Neuville & Péree, 1984) but provided conflicting evidence and the studies which had produced evidence for subject assignment used inter-sentential anaphora while the Maratsos study which did find evidence in support of parallel function used children and spoken language.

Therefore they decided to study single object pronouns in the second clause of conjoined sentences. Half the pronouns were ambiguous and half were made unambiguous by making the antecedents' gender different. The reading times were recorded and the assignments for the ambiguous sentences were inferred from questions following the presentation of each sentence. For example, (3.11) shows one of their ambiguous passages where the final clause of the final sentence was the target clause.

(3.11) Brenda and Harriet were starring in the local musical.
Bill was in it too and none of them were very sure of their lines or the dance steps.
Brenda copied Harriet and Bill watched her.

They found a subject assignment bias for the ambiguous sentences and the gender constraints facilitated comprehension in the unambiguous passages. Because of worries about subjects' memory for the antecedents' names needed for answering the questions and issues of immediacy, the experiment was repeated using only the ambiguous materials and subjects were asked to indicate their assignments immediately. The results showed a clear subject assignment bias and a possible facilitation of time-to-assign for subject antecedents.

Solan (1983) was concerned with what levels of representation parallelism operated over. He considered both grammatical and semantic parallelism because he observed that the two had been confounded in sentences like (3.12) where *burglar* has the same syntactic function (subject) and semantic function (agent) as the pronoun.

(3.12) The burglar saw the guard, and then he saw the police.

The study which he reports uses subjects from several age groups and concludes that young children use a semantic parallelism which is gradually replaced by a grammatical parallelism as adult grammar develops. One aspect to note is that only subject pronouns were used so the criticism of subject assignment and grammatical parallelism being confounded stands.

In summary, several investigations have tackled the issue of parallel function in sentence processing and in particular in pronoun resolution. These investigations have tried to disentangle the effects of parallel function from lexemic variables (causal bias), animacy, topic effects, subject assignment preferences, order of mention, level of syntactic representation, sentential connectives and markedness. Of the studies which investigated pronoun resolution only one (Crawley *et al.*, 1990) seriously attempted to separate subject effects from effects confounded with subjects (like topicality, givenness *etc.*) by using object pronouns as Caramazza and Gupta (1979) suggested which supported a subject assignment strategy. But because they omitted subject pronouns as well, they were unable to describe the subject assignment bias in relationship to the effects for subject pronouns. This point will become relevant when the experimental data is discussed. Therefore the best evidence argues against parallel function in pronoun resolution although Frazier *et al.*'s (1978) study did provide evidence for parallelism operating in general sentence processing. Of course, the evidence is not necessarily conclusive because a parallel function strategy may have been operating but simply been overridden by topic effects which is supported by Keenan's (1976) observation that subjects are often considered to be topics and controllers of anaphora. In spite of all these factors which have been considered in these studies only Cowan (1980) attempted to investigate the effects of order-of-mention. Given the work of Gernsbacher, Hargreaves and Beeman (1989) this seems to be an important area of consideration. Cowan claims that in his experiment assignments to conjoined subject antecedents were made at random which provided negative evidence for order-of-mention effects. However, out of the five sentence types he considered three showed a clear order-of-mention effect which he reported as not significant. Unfortunately, there is some doubt about his interpretation of the χ^2 test (Lewis & Burke, 1949). Therefore, the question of order-of-mention effects in pronoun resolution is largely unanswered.

Crawley *et al.* suggested that parallel function was used only as a "last resort" strategy when pragmatic information and other heuristics were inapplicable or when a sentence was particularly complicated. Stevenson (personal communication) had investigated sentences like (3.13) where general knowledge could have no influence and had found evidence for a parallel function strategy.

(3.13) The meter was next to the monitor and it was near the monitor.

These sentences were simple static descriptions and had little sense of temporal order to them so there was a possibility that with the addition of dynamism the effect of parallel function might be disrupted by other heuristics. For example pronouns often refer to salient items in a text (*e.g.*, Chafe, 1972; Sanford & Garrod, 1981) and the addition of dynamism might have increased the contrast in saliency between items. As Caramazza and Gupta (1979) noted passivisation might lead to an increase in the likelihood of a subject assignment strategy because the grammatical subject would be more closely identified with the theme/topic. The nouns used in Stevenson's study were all inanimate and as Stevenson (1979) observed animate subjects are given priority in a range of tasks. On the basis of these observations Experiment III was designed to investigate the effects of animacy and text style (dynamism was introduced by passivisation) on parallel function.

Crawley *et al.* had speculated on the causes of the subject assignment strategy which they had observed. They noted that all their antecedents were animate and that animate things are often agents (Dowty, 1991) and that animate subjects are often prioritised in psycholinguistic tasks (Stevenson, 1979). Therefore, Experiment IV was designed to test the effects of agency on parallel function.

A re-analysis of Crawley *et al.*'s data had suggested that there may have been an order-of-mention effect in their data. Some of the texts contained conjoined antecedents and others introduced their antecedents in a subject-predicate construction. Therefore Experiment V was designed to investigate this further by replicating their design while including a factor for the method of introducing the antecedents.

Inspection of Crawley *et al.*'s materials showed that in some of the materials the pronouns were not in fact grammatical objects but were part of prepositional phrases, for example. This meant that the pronouns should, perhaps, have been referred to as non-subject pronouns to cover all the categories in which they appeared. In the experiments reported in this chapter a similar situation arises because some of the experiments follow from Crawley *et al.*'s study. Therefore pronouns which are not grammatical subjects will be referred to as non-subject pronouns even when they are in fact grammatical objects, so as to make the naming conventions consistent.

Furthermore, order-of-mention and grammatical function or subjecthood *vs.* non-subjecthood is confounded because the first mentioned noun phrase in most clauses is a subject and the second mentioned noun phrase in most clauses is a non-subject. In the final experiment in this chapter this confound is addressed but in the other experiments it remains. Rather than continually mentioning the confound the first mentioned noun phrase will be referred to as a subject noun phrase and the second mentioned noun phrase will be referred to as a non-subject pronoun.

The first three experiments (Experiment III, Experiment IV and Experiment V) described in this chapter should be treated as pilot experiments. Their details are included here to make clear the design decisions taken for the subsequent experiments (Experiment VI, Experiment VII and Experiment VIII).

Animacy	Text	Example sentence
Animate	Dynamic	The girl was put next to the schoolmistress and she was pushed close to the schoolmistress.
Inanimate	Dynamic	The meter was put next to the monitor and it was pushed close to the monitor.
Animate	Static	The girl was next to the schoolmistress and she was close to the schoolmistress.
Inanimate	Static	The meter was next to the monitor and it was close to the monitor.

Table 3.1: Example sentence set for Experiment III. All the examples contain a pronoun in the subject position of the second clause and refer to an antecedent in the subject position of the first clause. The full set comprised sentences with all combinations of pronoun and antecedent position.

3.2 Experiment III

3.2.1 Subjects

The subjects were 48 students from Durham University who were volunteer participants in this experiment.

3.2.2 Design and Materials

There were four within subject design factors: Animacy (animate *vs.* inanimate), Text (Dynamic *vs.* Static spatial description), Pronoun position (subject *vs.* non-subject) and Antecedent position (subject *vs.* non-subject). Forty eight sentence sets were constructed and each set contained 16 sentences corresponding to all combinations of the four design factors applied to one of 48 *prototype sentences.* As an example, a reduced sentence set appears in Table 3.1 (examples showing the effects of the Pronoun and Antecedent variables have been left out).

The prototype sentences were generated from a set of 48 pairs of animate nouns, a set of 48 pairs of inanimate nouns and 48 verb phrase frames. A verb phrase frame comprised a pair of verbs and a pair of appropriate prepositional phrases. Each prototype sentence was made up from a pair of animate nouns, a pair of inanimate nouns and one verb phrase frame. No pair of nouns or verb phrase frame was repeated across the set of sentence prototypes. Within each prototype sentence the animate nouns were the same gender. Table B.1 shows the set of pairs of nouns. The verb phrase frames were generated by selecting at random 48 frames from the set of all possible frames generated from the verbs and prepositional phrases shown in Table 3.2. The final set of 48 verb phrase frames is shown in Table B.2.

All prototype sentences took the form of *The* N_1 *was* V_1 PP_1 *the* N_2 *and the* N_1 *was* V_2 PP_2 *the* N_2. The set of 16 sentences was then generated by taking combinations of elements from this sentence prototype and making appropriate substitutions which were determined by the levels of the four factors. Therefore, the Static sentences were generated by leaving out the Vs, the Animacy of the sentences was manipulated by choosing which pair of nouns to substitute

verbs: put, pushed, placed, moved, positioned, shoved
PPs: next to, beside, close to, near to, nearby, opposite to, to the left of, to the right of, behind, in front of, opposite, further away from

Table 3.2: Verbs and prepositions used to generate set of verb phrase frames from which to sample

and the antecedent of the pronoun was manipulated by choosing which noun to repeat. Of course, the Pronoun variable determined which N position in the second clause was replaced by a pronoun of the appropriate gender.

Two questions which took the form, *Was the N_1 PP N_2?* were paired with each sentence set. One question required a negative response to be correct (hereafter called a *false* question) and the other required a positive response to be correct (hereafter called a *true* question). Each verb frame had associated with it a particular noun order (*ab* or *ba*) which represented the order in which the nouns from the prototype sentence would appear in the questions. In half the question pairs the Ns were in the same order as the Ns in the final clause of the sentence (order ab) and half the time they were in the reverse order (order ba). A pair of prepositions used to form the questions was associated with each verb phrase frame. One preposition was used to form the true question and the other was used to form the false question. In order to stop subjects developing a matching strategy the prepositions in the questions were not always the same as the prepositions which appeared in the sentence. As an example, consider the formation of the two questions associated with the first verb phrase frame taken from Table B.2. The final clause of the sentence in the Dynamic condition would have been of the form *and the N_1 was shoved opposite to the N_2*. The noun order for the verb phrase frame is *ab* which means that the order of the nouns in the question was N_1 followed by N_2. Had the order been *ba* then N_2 would have followed N_1 in the questions. The true preposition was *facing* so the true question was, "Was the N_1 facing the N_2?" and the false preposition was *beside* so the false question was "Was the N_1 beside the N_2?".

Sixteen material sets were formed by assigning one of the 16 versions of each sentence set to each material set. Therefore across the 16 material sets each prototype sentence occurred in all 16 of its versions. However, within each material set, each prototype sentence occurred in one of its 16 versions. Therefore, there were 3 replications (48 divided by 16) of each condition in any one material set. Every fourth sentence in a material set was followed by a question so that 12 (48 divided by 4) of the possible 16 conditions in each material set was followed by a question. The questioning was balanced so that each condition was questioned the same number of times across the 16 material sets. Half the questions required a negative response and half required a positive response to be correct. Three subjects were randomly assigned to each material set, ensuring that each subject was only exposed to one version of a sentence set. The order of the sentences and therefore the interval between questions, was randomized.

3.2.3 Procedure

The task was a self-paced reading task. Each sentence was presented as two clauses, one at a time, on the screen of a BBC microcomputer. Subjects were asked to press the space bar

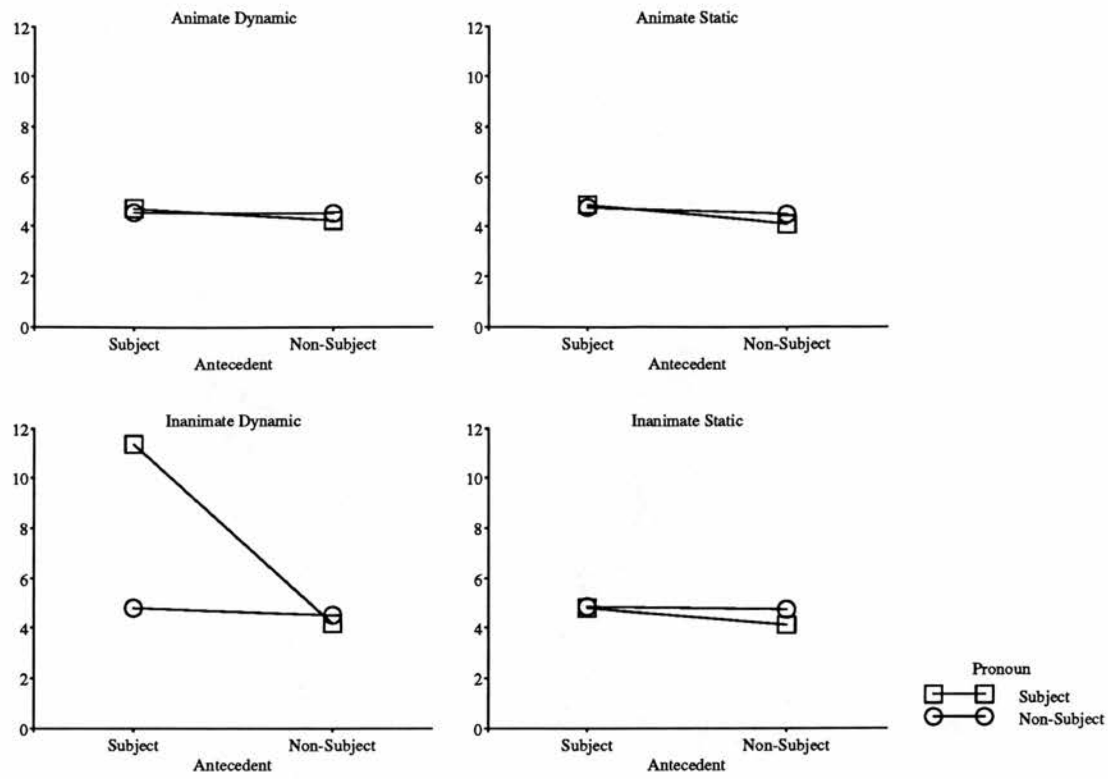

Figure 3.1: Reading rates in words per second by Animacy (Animate and Inanimate), Text type (Dynamic and Static), Pronoun (Subject and Non-subject) and Antecedent (Subject and Non-subject) for Experiment III.

as soon as they had read and understood the clause. The next clause was then presented. When the second clause had been read the screen cleared and on appropriate trials a question appeared. After answering the question by pressing one of two keys marked *true* and *false* subjects were prompted to start the next trial. If there was no question the prompt for the next trial appeared immediately.

The time taken to read each target (second) clause was measured in milliseconds and the answer to each question was recorded.

3.2.4 Results

Outliers were considered to be times less then 350ms and times greater than 10000ms. They were were replaced by their cut-off values, so a time of 200ms, for example, was replaced by 350ms. There were 2304 data, 50 (2.2%) fell below the lower cut-off (mean number of outliers per subject=1.042, SD=0.202) and 15 (0.7%) fell above the upper cut-off (mean number of outliers per subject=0.312, SD=1.075).

Clearly, Dynamic sentences contain more words than corresponding Static sentences because the Dynamic sentences contained a complex verb. Therefore the reading times for each clause were converted to reading rates by dividing the number of words by the reading time and multiplying by 1000 to produce a rate in words per second.

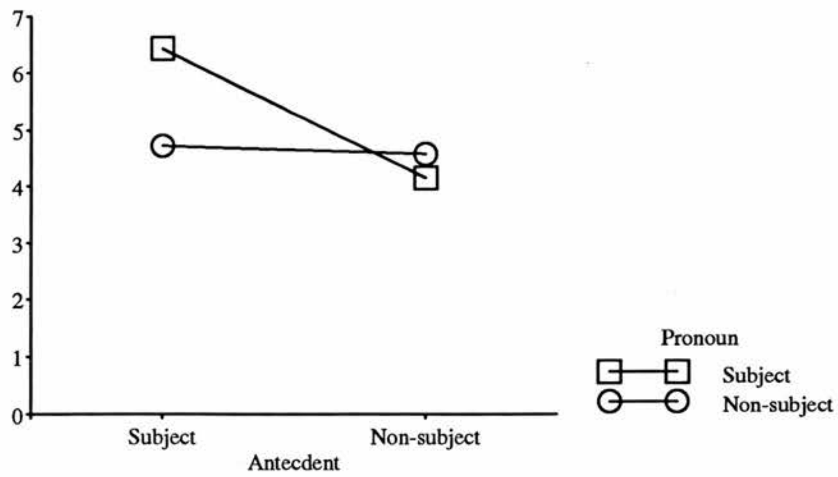

Figure 3.2: Reading rates in words per second by Antecedent (Subject and Non-subject) and Pronoun (Subject and Object) for Experiment III.

Figure 3.1 shows the reading rates in all conditions averaged over subjects. It is clear from this table that there is a large facilitatory effect for subject pronouns which refer to subject antecedents in Dynamic sentences about inanimate objects. Analysis of variance confirms that this interaction (between Animacy, Text, Pronoun and Antecedent) is reliable ($minF'(1,58) = 17.08, p < 0.01$). All the lower order interactions and effects are also reliable but it is clear that almost all of them are artificial in the sense that they result from the single facilitatory effect described above. However there does appear to be a genuine interaction between antecedent position and pronoun position: non-subject pronouns are unaffected by their antecedent's position and clauses with subject pronouns are facilitated if they refer to antecedents in subject position. Figure 3.2 shows the means for the interaction. Table 3.3 shows simple main-effect tests for the effect of antecedent position on subject and non-subject pronouns at the four conditions of Animacy by Text. The tests lend support to the observation that non-subject pronouns are unaffected by the position of the antecedent whereas subject pronouns are processed faster when they refer to an antecedent in the subject position of the preceding clause. Of course, the interaction could be read the other way in which pronoun position had an effect on antecedent position: subject antecedents were facilitated by subject pronouns and non-subject antecedents were facilitated by non-subject pronouns. This reading sounds much more like the anticipated parallel function interaction but of course if the predicted parallel function effect was really there then both readings of the interaction would be consistent.

Figure 3.3 shows the mean reaction times by condition and the pattern clearly replicates the pattern of reading rates in Figure 3.1. Furthermore Analysis of Variance shows that there is a reliable four-way interaction between Animacy, Text, Antecedent and Pronoun ($minF'(1,92) = 4.15, p < 0.01$). This analysis of the reading times supports the use of reading rates and the transformation. Analysis of the question responses showed that 59% of the answers were incorrect and there did not appear to be any effect of response bias (62% of questions which required positive response were answered incorrectly and 55% of the questions which required a negative response were answered incorrectly). Table B.8 shows the number of correct and incorrect responses by antecedent and pronoun position. Table B.9

Text	Animacy	Pronoun	dfs	F_1	$p<$	F_2	$p<$	dfs	$minF'$	p
Narrative	Animate	Subject	1,47	4.21	0.05	1.81	0.19	1,81	1.27	ns
Narrative	Animate	Non-subj	1,47	0.42	0.52	0.14	0.71	1,75	0.10	ns
Narrative	Inanimate	Subject	1,47	60.45	0.01	32.16	0.01	1,86	20.99	<0.01
Narrative	Inanimate	Non-subj	1,47	1.98	0.17	0.64	0.43	1,74	0.48	ns
Spatial	Animate	Subject	1,47	10.64	0.01	11.13	0.01	1,94	5.44	<0.05
Spatial	Animate	Non-subj	1,47	2.06	0.16	0.81	0.37	1,79	0.58	ns
Spatial	Inanimate	Subject	1,47	11.62	0.01	9.4	0.01	1,93	5.20	<0.05
Spatial	Inanimate	Non-subj	1,47	0.74	0.40	0.26	0.61	1,76	0.19	ns

Table 3.3: Simple interactions by subjects and materials analysis by text, animacy and pronoun position. In all cases the actual vaule of p is smaller than stated. The Non-subject level of Pronoun has been abbreviated to Non-subj.

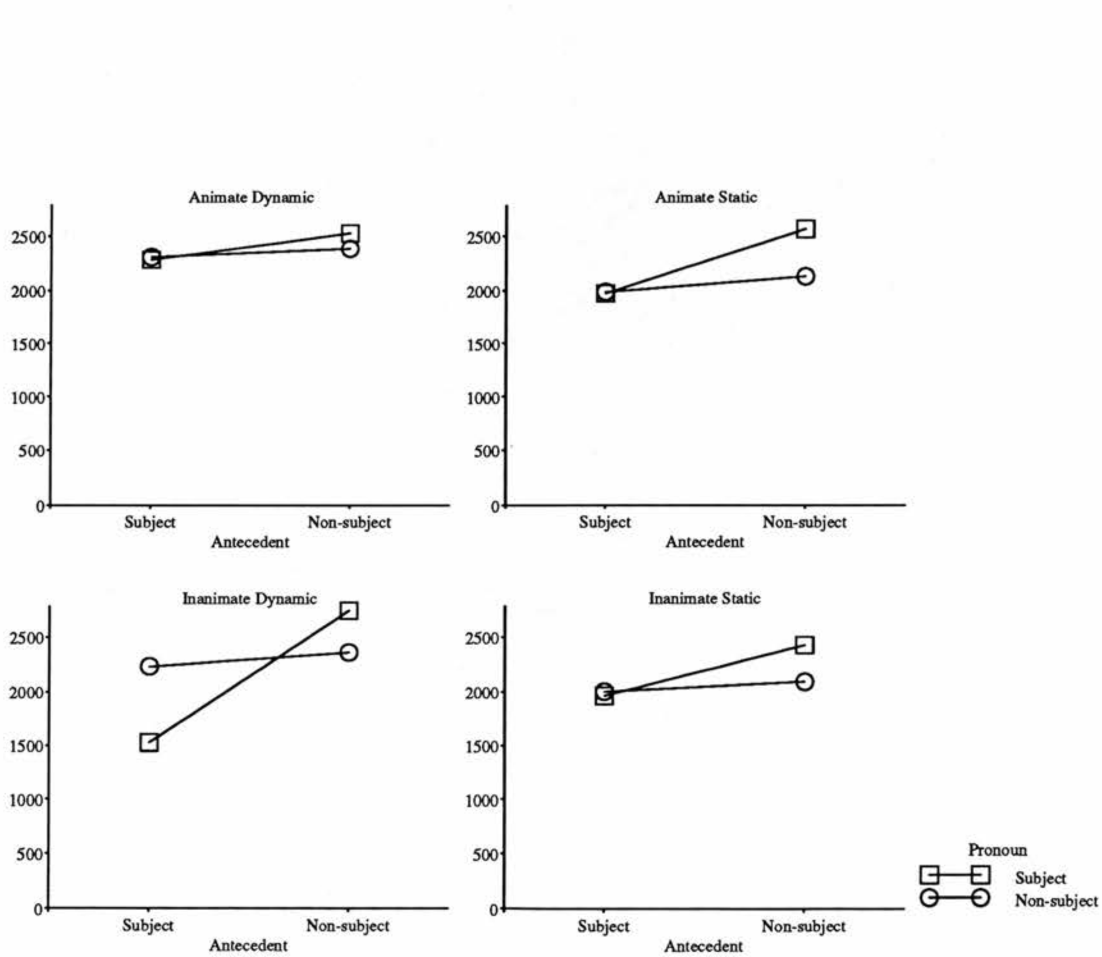

Figure 3.3: Mean reading times (milliseconds) by Animacy (Animate and Inanimate), Text type (Dynamic and Spatial), Pronoun (Subject and Non-subject) and Antecedent (Subject and Non-subject) for Experiment III.

shows the number of correct and incorrect responses to questions by all four factors. The tables suggests that different conditions are easier to comprehend than others. However, no analysis is possible because the subjects and materials were not questioned an equal number of times across conditions.

3.2.5 Discussion

The results show that the second clause of sentences like *The meter was put next to the monitor and it was pushed close to the monitor* are very quick to read. There is a general result that the processing of subject pronouns is facilitated by subject antecedents and the second mentioned pronouns (in prepositional phrase) show no effect of antecedent position. This general effect has two explanations. First, there could be two strategies operating in competition: subject assignment and parallel function. For subject pronouns the two strategies agree and so an effect would be observed and for non-subject pronouns the two heuristics predict opposite results which could cancel out producing no effect. Second, subject pronouns are ambiguous when encountered whereas non-subject pronouns are not because the repeated noun phrase comes before them. Therefore the lack of an effect of antecedent position for non-subject pronouns may simply be a sign that they are unambiguous and so no reassignment is necessary. The results cannot discriminate between these explanations and certainly do not discriminate between a parallel function strategy or a subject assignment strategy.

3.3 Experiment IV

The aim of this experiment was to investigate the effects of agency on any parallelism effects in pronoun resolution.

3.3.1 Subjects

The subjects were 48 volunteer students from Durham University.

3.3.2 Materials and design

The design factors were: Agency (agent *vs.* non-agent), Pronoun position (subject *vs.* non-subject) and Antecedent position (subject *vs.* non-subject). Thirty two sentence sets were constructed each of which contained 8 sentences corresponding to all combinations of the three design factors applied to a prototype sentence. Table 3.4 shows an example sentence set. The prototype sentences were generated from a set of 32 agentive verbs and associated adverbial phrase (see Table B.6), 32 non-agentive nouns and associated adverbial phrase and 32 animate nouns (see Table B.5). All the sentences in a sentence set were of the form *The* N_1 VP_1 *the* N_2 *and the* N_1 VP_2 *the* N_2. One of the N positions in the second clause was occupied by a pronoun depending on the value of the Pronoun position variable and the index of the remaining N depended on the level of the antecedent variable.

Agency	Example sentence
Agent	The schoolmistress punished the girl and she apologised to the girl much later.
Non-agent	The schoolmistress heard the girl and she surprised the girl by being there.

Table 3.4: Example sentence set for Experiment IV. Only examples for subject pronouns with subject antecedents are shown.

To ensure subjects' comprehension each prototype sentence was followed by a question about the individual's state (*e.g.*, Was the schoolmistress/girl sorry for what she had done?). Half the questions required a negative answer to be correct and half the questions required a positive answer.

Eight material sets of 48 sentences each were formed by assigning one of the 8 versions of each sentence set to each material set. Therefore across the 8 material sets each prototype sentence occurred in all 8 of its versions. However, within each material set, each prototype sentence occurred in one of its 8 versions. The were 4 replications of each condition in any one material set. Every fourth sentence in a material set was followed by a question so that every condition was questioned once and 4 conditions twice. The questioning was balanced so that each condition was questioned the same number of times across the 16 material sets. Half the questions required a negative response and half required a positive response to be correct. Forty-eight subjects were randomly assigned to each material set, ensuring that each subject was only exposed to one version of a sentence set. The order of the sentences and therefore the interval between questions, was randomized.

3.3.3 Procedure

The procedure was the same as in Experiment III.

3.3.4 Results

Outliers were treated in a similar way to those in Experiment III. Out of 1536 data, 1 (0.1%) datum fell below 350ms (mean number per subject=0.021,SD=0.144) and 29 (1.9%) data above 10000ms (mean number per subject=0.604,SD=2.430). Figure 3.4 shows that clauses with agentive nouns were read faster than clauses with non-agentive nouns. Analysis of variance showed that this difference was not reliable ($F_1(1,47) = 7.66, p < 0.009$, $F_2(1,31) = 3.83, p < 0.06$). There was also an interaction between pronoun and antecedent position: subject pronouns are read faster with subject antecedents and non-subject pronouns are unaffected by antecedent position ($minF'(1,78) = 3.98, p < 0.05$), Figure 3.5 shows the means.

Analysis of the question responses showed that 39% of the answers were incorrect and there did appear to be a response bias (24% of the questions which required a positive response were given a negative response and 55% of the questions which required a negative response were given a positive response). Table B.12 shows that it is easier to get the question right

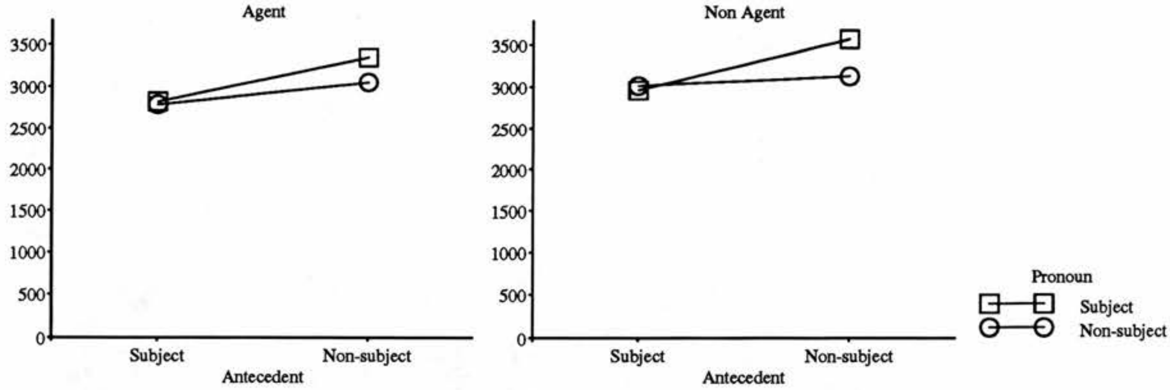

Figure 3.4: Mean reading times (in milliseconds) for the second clause in Experiment IV by Agency (Agent and Non Agent), Antecedent (Subject and Non-subject) and Pronoun (Subject and Non-subject).

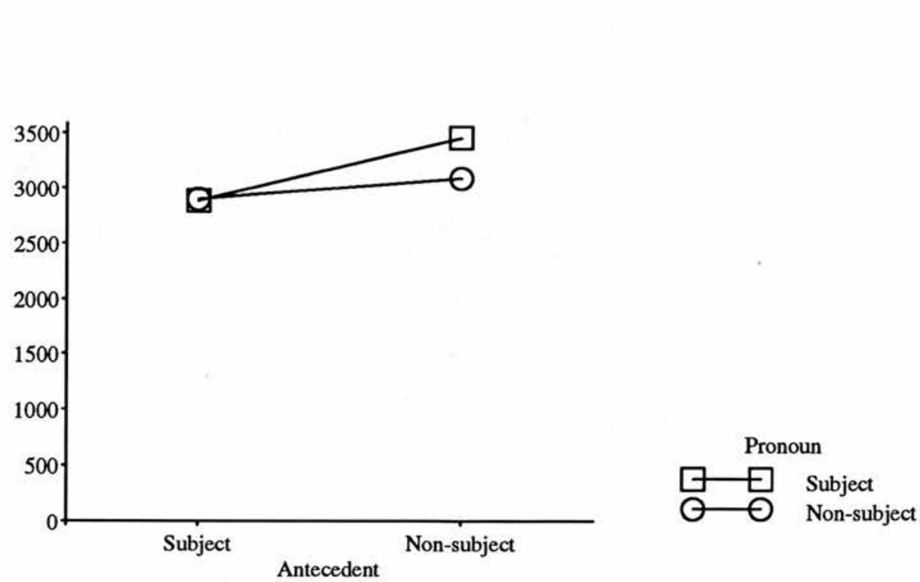

Figure 3.5: Mean reading times (in milliseconds) by Antecedent (Subject and Non-subject) and Pronoun (Subject and Non-subject) in Experiment IV.

for agentive sentences than for non-agentive sentences. However for reasons similar to those in Experiment III no confirmatory statistics can be reported. Table B.13 shows that there is very little effect of pronoun and antecedent position on questions response accuracy.

3.3.5 Discussion

Although the effect of agency was not reliable it was very close. This result supports the observation that pronouns commonly refer to agents rather than non-agents (Dowty, 1991) and would therefore be expected to be processed more easily. The interaction between pronoun position and antecedent position replicates the interaction found in Experiment III. Subject pronouns do show an effect of antecedent position and non-subject pronouns do not. This experiment has confirmed that a pronoun is easier to process if its antecedent is an agent rather than a non-agent.

3.4 Experiment V

This experiment was essentially a replication of Crawley *et al.*, (1990) with the inclusion of a factor for the pattern of antecedent introduction in the first sentence of the texts.

3.4.1 Subjects

The subjects were 48 volunteer students from Durham University.

3.4.2 Materials and Design

Each subject read 48 passages consisting of two context sentences followed by a target sentence. The design factors were text type (conjoined antecedents, subject predicate antecedents and separate antecedents) and participant order (same order *vs.* different order). Half the targets contained 3 or 4 words after the *and* and half the targets contained 5 to 7 words after the *and* making a within materials variable of length with two levels. Table 3.5 shows a reduced text set illustrating the three levels of the text style variable: conjoined antecedents, subject predicate antecedents and separate antecedents. Another variable, namely Order of Participants was manipulated: either the order of mention of the participants was the same in context and target or the order was different. The pronoun in the target sentence was always ambiguous. Each text was followed by a question derived from the crucial part of the second clause in the target sentence by repeating the second clause with the pronoun replaced by one of the potential antecedents (*e.g.*, Ellen kicked Sammy? or Ellen kicked John?). The question was used to determine the assignment of the pronoun in the preceding text. The number of times each potential antecedent was substituted was balanced across materials.

Forty-eight text sets were constructed. Each set contained 6 texts corresponding to all combinations of the two design factors applied to a target sentence. See Table B.7. Six material sets

Example passage
Conjoined antecedents
John and Sammy were playing in the garden.
Ellen watched their game with interest.
John pushed Sammy and Ellen kicked him.
Antecedents in subject predicate form
John was playing with Sammy in the garden.
Ellen watched their game with interest.
John pushed Sammy and Ellen kicked him.
Antecedents in separate senteces
John and Ellen were playing in the garden.
Sammy watched their game with interest.
John pushed Sammy and Ellen kicked him.

Table 3.5: A target sentence with three different context sentence pairs.

were formed by assigning one of the 6 versions of each text set to each material set. Therefore across the 6 material sets each prototype sentence occurred in all 6 of its versions. However, within each material set, each prototype sentence occurred in one of its 6 versions. There were 8 replications of each condition in any one material set. Forty-eight subjects were randomly assigned to each material set, ensuring that each subject was only exposed to one version of a text set. The order of the texts was randomized.

3.4.3 Procedure

The task was a self-paced reading task. The context sentences of the passages were presented one sentence at a time and the target sentence was presented clause by clause. Subjects were asked to press the space bar as soon as they had read and understood the sentence/clause. Once a sentence appeared on the screen it stayed there until the passage was complete. Once the final clause had been read, the screen cleared and then the question appeared. After answering the question by pressing one of two keys marked *true* and *false*, subjects were prompted to start the next trial.

The time taken to read the last clause of each target sentence was measured in milliseconds and the answer to each question was recorded.

3.4.4 Results

Outliers were treated in a similar way to those in Experiment III. Out of 2304 data, 3 (0.1%) data fell below 350ms (mean number per subject=0.062,SD=0.245) and 17 (0.7%) above 10000ms (mean number per subject=0.354,SD=0.838).

First an analysis was carried out to determine whether there was any response bias. For each subject, for each of the two possible responses the number of non-subject assignments was subtracted from the number of subject assignments. Analysis of variance showed that there was no effect of response on assignment (both Fs < 1).

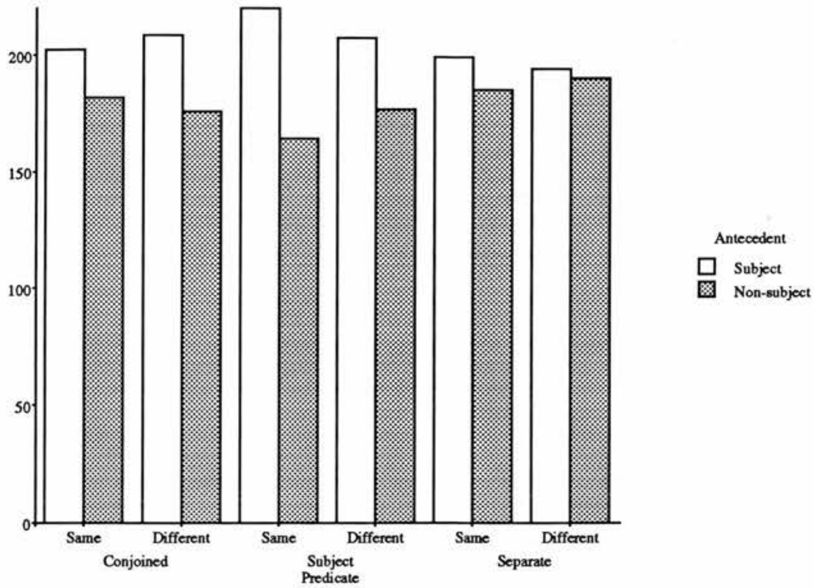

Figure 3.6: The number of assignments to subject and object antecedents in Experiment V by Text (Conjoined, Subject-Predicate and Separate) and Order (Same and Different) where N=384 for each condition.

Figure 3.6 shows the number of assignments made to the subject and non-subject antecedents by the three types of text and by the order of the antecedents. In order to prepare the data for analysis of variance the number of non-subject assignments was subtracted from the number of subject assignments for each condition for each subject. Analysis of variance (subject and materials analysis) showed that there were no significant effects of Text or Antecedent order. For Text, $F_1(1,47) = 2.45, p < 0.1$, $F_2(1,47) = 1.69, p < 0.2$ and for all the other comparisons $F < 1$. There is a clear subject assignment strategy across all conditions. However the size of it varies across the three text types. The effect is largest in the Subject-Predicate condition, slightly smaller in the Conjoined condition and smaller still in the Separate condition. Figure 3.7 shows the transformed data by Text and Order. Figure 3.8 shows the mean reading rates by condition. Rates were used rather than raw RTs to remove any effects of clause length. Analysis of variance by subject and materials is not possible because there are empty cells. Maximum Likelihood analysis of the reading time data with Text, Order, Assignment and Length only shows a significant effect of Length ($F(1,2280) = 29.13, p < 0.0001$) and nothing else. This analysis shows that subjects were reading the target sentences because the effect of the number of words in a target was in the expected direction: the more more words the longer the target took to read.

3.4.5 Discussion

The assignment results show that there is a clear bias to assign pronouns to the first mentioned antecedent or the antecedent in the subject position of the last sentence. This could simply be a subject assignment strategy over the target sentence or it could be related to the context sentences as well. The assignment data imply that Text and Order do matter although the interaction is not reliable and this means that the assignment bias operates over the whole

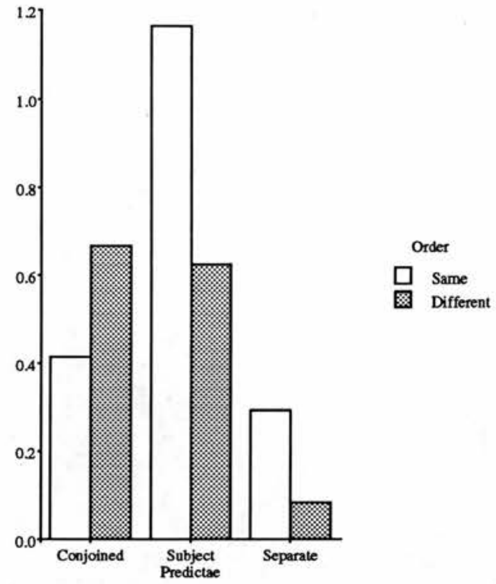

Figure 3.7: Pronoun assignment score in Experiment V by Order (Same and Different) and Text (Conjoined, Subject-Predictae and Separate) where N=48 for each condition.

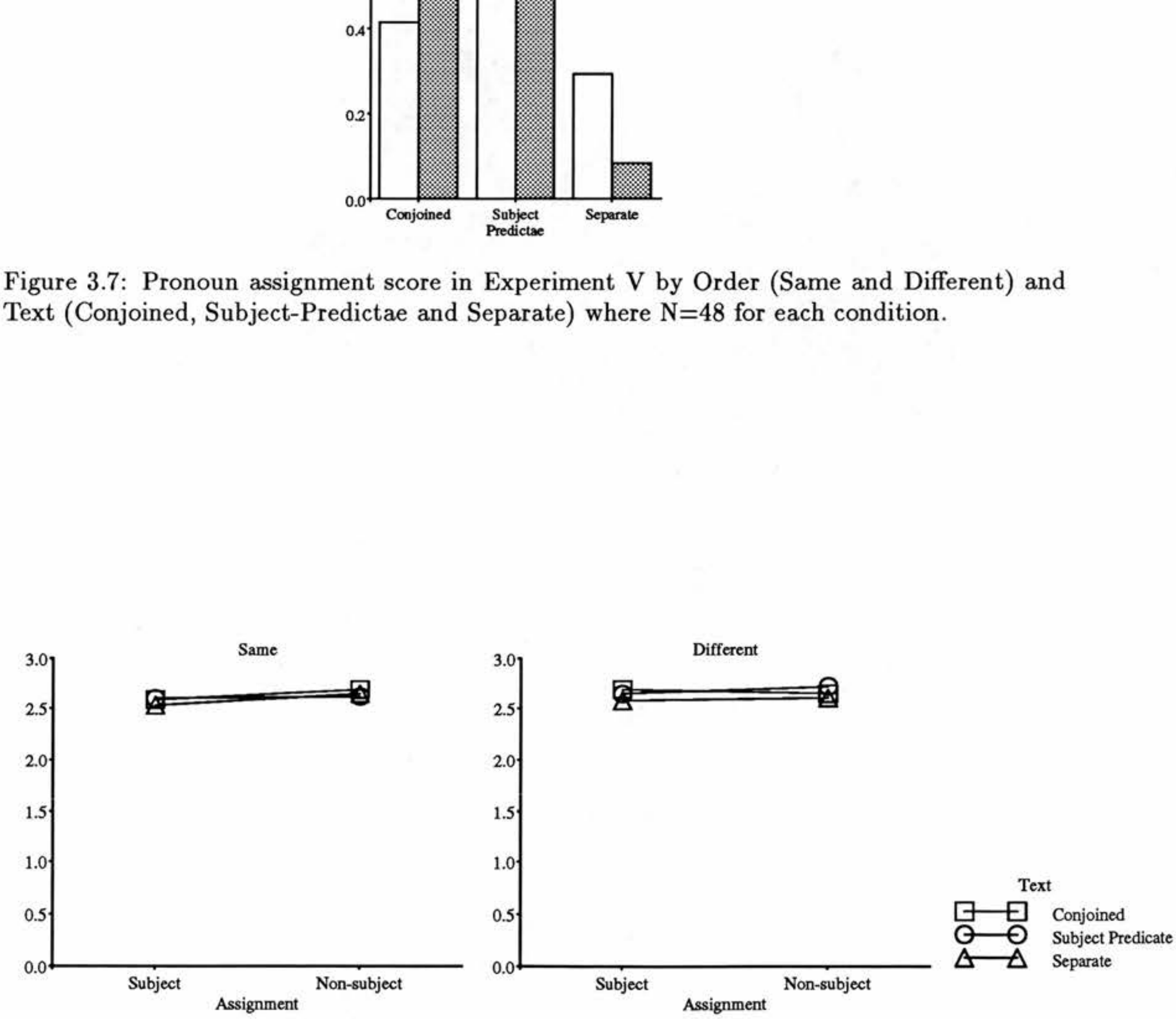

Figure 3.8: Reading rates in words per second by Text (Conjoined, Subject Predicate and Separate), Assignment (Subject and Object) and Order (Same and Different) in Experiment V.

text and not just the last sentence.

However, these results are difficult to asses because of three aspects of the design. First, there are no subject pronouns so it is not possible to determine if there is a parallel function effect. Second, there is no condition where only the last sentence is presented alone so as to have a baseline measure of any assignments or reading time biases and to asses the effects of the context sentences. Thirdly, the Text variable is a within variable which may cause strategic confusions and leads to a small amount of data (some cells are empty).

Although it is difficult to conclude anything about parallel function or Text style from these results they do hint that at least Text style may have an effect on pronoun assignment. This implies that the effects of ordering among antecedents can operate across sentences.

3.5 General Discussion of Experiments III, IV and V

Experiments III and IV both show similar general results with regard to the question of parallel function. Pronouns which are in the subject position of the target clause do show an effect of antecedent position: clauses which contain pronouns which have antecedents in the subject position are read faster than those with have antecedents in the non-subject position of the previous clause. This result is consistent with a parallel function heuristic, with a subject assignment heuristic and an advantage of first mention effect. Non-Subject pronouns should be able to separate two of the possibilities but these two experiments show that there is no effect of antecedent position and this result can be explained by two possible explanations.

First, consider the following sentence *The girl was put next to the schoolmistress and she was pushed close to the schoolmistress*. When the pronoun is encountered there is nothing to tell the reader which potential antecedent it refers to. Therefore, the reader may well make an initial assignment based on some heuristic and proceed through the rest of the clause. When the second noun phrase is encountered the pronoun is no longer ambiguous and depending on the initial assignment the pronoun may need to re-assigned which may well take extra time. Presumably it is this sort of process which is causing the difference between antecedents for subject pronouns.

However the process is probably different for sentences like *The girl was put next to the schoolmistress and the schoolmistress was pushed close to her*. Here there is no point at which the pronoun might need to be re-assigned. Once the first noun phrase of the second clause is encountered the pronoun's referent is fully determined and it is therefore not surprising that there is no effect of antecedent position because there is no need for re-assignment in any condition.

This difference between subject and non-subject pronouns depends on the experimental subjects assuming that there will always be a pronoun in the second clause which refers to one of the two antecedents referred to in the first clause and that one of the two antecedents will be mentioned in the second clause, thus disambiguating the pronoun.

The second explanation for the lack of effect of antecedent position on non-subject pronouns supposes that both parallel function and first-mention/subject assignment are operating. For subject pronouns the two strategies agree on assignment to the subject antecedent. For non-subject pronouns the predictions of the two strategies disagree: the parallel function strategy predicts assignment to the non-subject antecedent and the subject assignment strategy predicts assignment to the subject. If the processes are averaged over in these experiments then the two strategies may cancel out showing no effect of antecedent position for sentences with non-subject pronouns.

These experiments provide little information about parallel function because the non-subject pronouns were unambiguous and are essentially pilot experiments. Therefore, in order to investigate the effects of text style and animacy on parallelism, Experiment III was repeated in two versions without repeating either of the first-clause noun phrases. In one version, the pronouns were left ambiguous and subjects were asked to report which antecedent they thought the pronoun referred to. In the second version the pronouns were disambiguated by using gender and number which meant that the pronouns were ambiguous up until they were read so that non-subject pronouns did not suffer the same problems as in Experiment III.

Experiment V was designed to examine the effects of context on sentence level heuristics and investigate order of mention. There is a hint in the assignment data that there are effects which carry over from the context into the target sentence but the results are not reliable. Also there are no subject pronouns in the materials so it is not possible to differentiate between a subject assignment strategy and a parallel function strategy. There may also be problems with making Text style a within-subjects factor because subjects may not be able to develop a consistent strategy for dealing with the texts and it also means that there is very little data per subject. Therefore another version of Experiment V was conducted which included a baseline condition in which only the target sentence was presented, Text style was made a between-subject factor and subject pronouns were included.

3.6 Experiment VI

The aim of this experiment was to repeat Experiment III with the addition of a third individual in the target clause to try to make the non-subject pronouns equally ambiguous with the subject pronouns.

3.6.1 Subjects

The subjects were 48 students from Durham University.

3.6.2 Design and Materials

There were three design factors: Animacy (animate *vs.* inanimate), Text (Dynamic *vs.* Static description) and Pronoun position (subject *vs.* non-subject). Forty-eight sentence sets were constructed and each set contained 8 sentences corresponding to all combinations of the three design factors applied to a prototype sentence. As an example, a reduced sentence set appears in Table 3.6. The prototype sentences were generated from a set of 48 triples of animate nouns

Animacy	Text	Example sentence
Animate	Narrative	The girl was put next to the schoolmistress and she was pushed close to the prefect.
Animate	Spatial	The girl was next to the schoolmistress and she was close to the prefect.
Inanimate	Narrative	The meter was put next to the monitor and it was pushed close to the dial.
Inanimate	Spatial	The meter was next to the monitor and it was close to the dial.

Table 3.6: Example sentence set for Experiment VI. All the examples contain an ambiguous pronoun in the subject position of the second clause. The full set comprised sentences with pronouns in the object position of the second clause.

and a set of 48 triples of inanimate nouns and 48 verb phrase frames. A verb phrase frame comprised a pair of verbs and appropriate prepositional phrases. Each prototype sentence was made up from three animate nouns, three inanimate nouns and one verb phrase frame. No triple of nouns or verb phrase frame was repeated within a sentence prototype or across the set of sentence prototypes. Table B.3 shows the set of triples of nouns. The verb phrase frames were the same as in Experiment III (see Table B.2).

All prototype sentences took the form of *The* N_1 *was* V_1 PP_1 *the* N_2 *and the [PRO or* N_3*] was* V_2 PP_2 *the [PRO or* N_3*]*. The set of 8 sentences was then generated by taking combinations of elements from this sentence prototype and making appropriate substitutions which were determined by the levels of the three factors. Therefore, the Static sentences were generated by leaving out the Vs, the Animacy of the sentences was manipulated by choosing which triple of nouns to substitute (in the animate sentences the nouns which occurred in the first clause were always the same gender). The Pronoun variable determined which N position was replaced by a pronoun of the appropriate gender.

Eight material sets were formed by assigning one of the 8 versions of each sentence set to each material set. Therefore across the 8 material sets each prototype sentence occurred in all 8 of its versions. However, within each material set, each prototype sentence occurred in one of its 8 versions. The were 6 replications of each condition in any one material set. Six subjects were randomly assigned to each material set, ensuring that each subject was only exposed to one version of a sentence set. The order of the sentences was randomized. The material lists were printed on small booklets made from stapled sheets of paper.

3.6.3 Procedure

Each subject was given a booklet and asked to read the sentences and treat each one separately. After reading each sentence the subjects were required to underline the object or individual to which they thought the pronoun in the sentence referred. Subjects were asked not to spend too much time on each one.

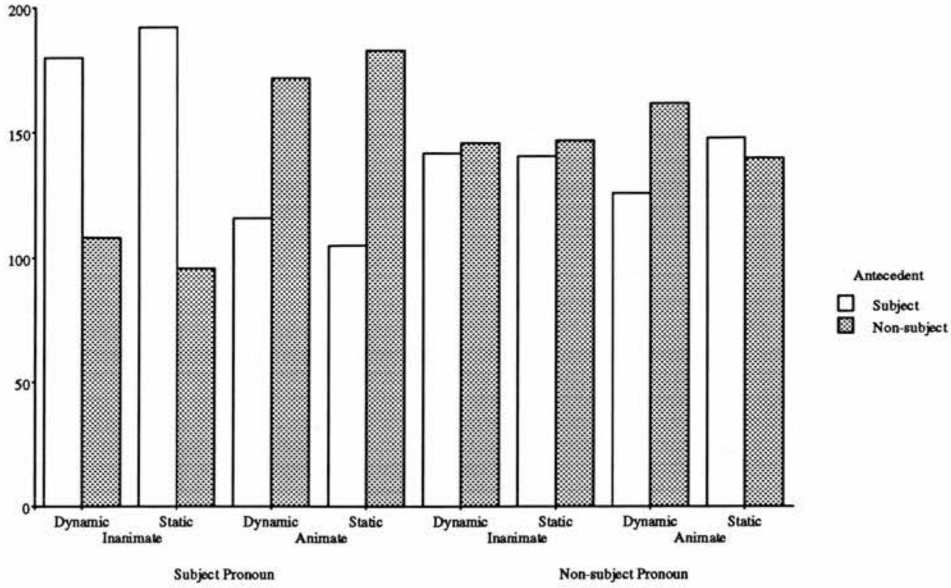

Figure 3.9: Number of pronoun assignments to subject and non-subject antecedents by Text (Dynamic and Static), Animacy (Inanimate and Animate) and Pronoun (Subject and Non-subject) in Experiment VI where N=288 for each condition.

3.6.4 Results

Figure 3.9 shows the number of subject and non-subject pronoun assignments by Text style, Animacy and Pronoun position. To prepare the data for Analysis of variance the number of non-subject assignments was subtracted from the number of subject assignments for each condition for each subject to give a single dependent variable. Figure 3.10 shows the transformed data. Figure 3.10 shows that in sentences about inanimate things, subject pronouns are likely to be assigned to subject antecedents and non-subject pronouns to non-subject antecedents. However, the size of the effect for non-subject pronouns is small compared to the size of the effect for subject pronouns. In fact the results are better stated thus: subject pronouns are likely to be assigned to subject antecedents and non-subjects pronouns show only a small bias to non-subject antecedents over subject antecedents. The style of sentence (Static or Dynamic) does have a small effect: Static sentences seem to increase the biases produced in assignments by the pronoun position.

Sentences about animate things show a very different pattern. Subject pronouns are more likely to be assigned to non-subject antecedents than to subject antecedents. Again the size of this effect seems to be increased in Static sentences. Non-Subject pronouns in Dynamic sentences show a bias to non-subject antecedents whereas non-subject pronouns in Static sentences show a bias to subject antecedents.

In summary, subject pronouns in inanimate sentences show a bias in assignments to subject antecedents whereas in animate sentences they show a bias to non-subject pronouns. Non-Subject pronouns have the effect of reducing the bias shown by subject pronouns. Text style has a small effect: Static sentences seem to increase the size of any other effects compared to Dynamic sentences. Analysis of variance (on the transformed data) confirmed that there was an effect of Animacy ($minF'(1,86) = 14.30, p < 0.01$, 0.82 for Inanimate and -0.84

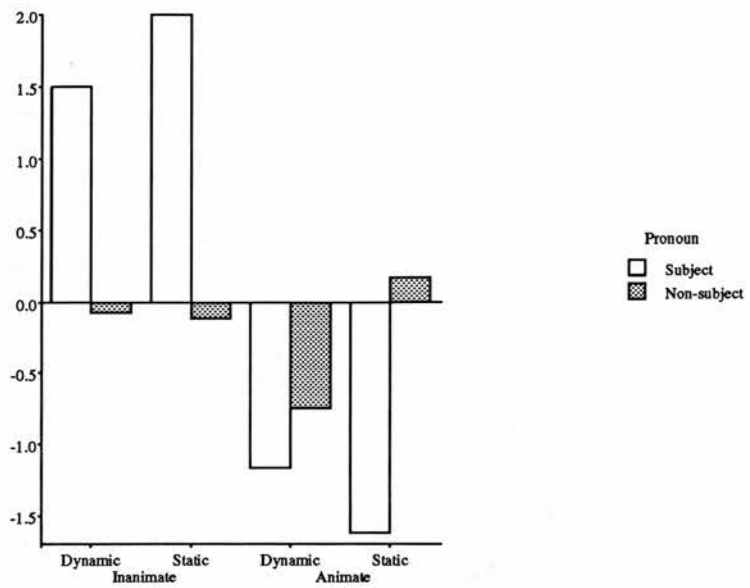

Figure 3.10: Pronoun assignment score by Animacy (Animate and Inanimate), Text (Conjoined, Subject-Predicate, Separate and Single) and Pronoun (Subject and Non-subject) in Experiment VI where N=48.

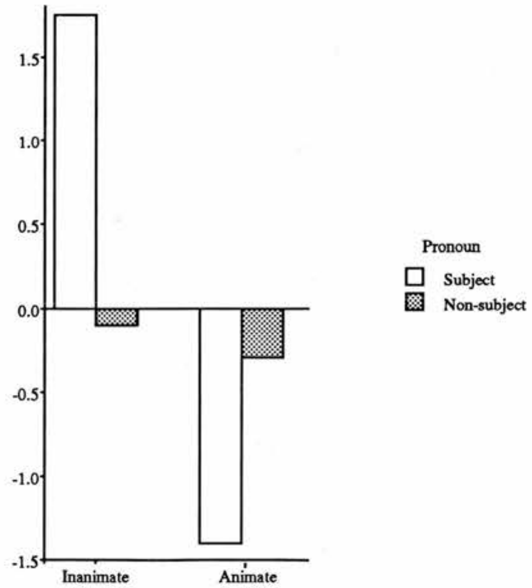

Figure 3.11: Pronoun assignment score by Animacy (Inanimate and Animate) and Pronoun (Subject and Non-subject) in Experiment VI.

for Animate) and an interaction between Pronoun position and Animacy ($minF'(1,85) = 16.73, p < 0.01$, see Table 3.11). The interaction between Pronoun position, Animacy and Text style approaches significance, $F_1(1,47) = 2.94, p < 0.1$ and $F_2(1,47) = 3.40, p < 0.08$. A simple effects analysis on the non-subject level of the Pronoun factor showed that Text was not significant ($F < 1$), Animacy had a main effect ($minF'(1,86) = 8.43, p < 0.05$) and that there was an interaction between Text and Animacy ($minF'(1,94) = 13.83, p < 0.05$). This last interaction shows that at least the bias for non-subject assignments shown by non-subject pronouns in the Animate Dynamic condition is different from non-subject pronouns behaviour in the other three conditions. Crucially, the interaction between Text and Pronoun is not significant at either the Inanimate level (F_1 and $F_2 < 1$) or the Animate level ($F_1(1,47) = 3.13, p < 0.09, F_2(1,47) = 2.77, p < 0.2$).

3.6.5 Discussion

There seems to be a clear subject assignment strategy for subject pronouns in inanimate sentences. This bias is severely reduced in the assignments of non-subject pronouns. Curiously, subject pronouns exhibit a reverse bias in animate sentences but non-subject pronouns still behave the same way in that they reduce that bias. Clearly subject pronouns exhibit strong biases which are affected by the animacy of the things to which they refer. However, non-subject pronouns are either insensitive to these biases or in fact counteract their effect to some extent. Although this effect does seem clear it does in fact conflict with the results from later experiments. This conflict will be returned to in Section 3.9.3 and a possible explanation for these particular materials will be offered here.

None of the results exhibit a "strict" parallel function effect, in that, the size of the non-subject pronoun biases are not of the same size as the subject pronoun biases. First consider the results for the inanimate materials because they most clearly conform to the general predictions for this experiment. There are two possible reasons for a small effect of the non-subject pronouns. First, non-subject pronouns could simply be different from subject pronouns. Experiments designed to investigate pronoun resolution very rarely use non-subject pronouns and the results which most reliably apply to pronouns only apply to subject pronouns. Therefore, the subject pronouns in this experiment may well be resolved using a subject assignment strategy or a parallel function strategy: the two strategies mean the same thing if only applied to subject pronouns. The non-subject pronouns show a very small and probably insignificant bias which may be because they are being assigned at random because heuristics are not used to help in their resolution.

The second reason could be that there are two heuristics operating simultaneously: a subject assignment strategy and a parallel function strategy. When a subject pronoun is resolved the two strategies will agree and the pronoun is likely to be assigned to the antecedent in the subject position. However, when a non-subject pronoun is encountered the two strategies disagree and the choice between the two antecedents may be made at random. Again, if the choice was made at random then no bias would be expected which agrees with the results.

The results for the animate materials are quite different from the inanimate materials and require a different explanation although they show the same pattern of strong bias for subject pronouns and very little bias for non-subject pronouns. Clearly, neither a subject assignment heuristic nor a parallel function heuristic can be operating. One possible explanation for these materials may be related to recency. Suppose that animate things cause the processor considerable difficulty in this task. Goodluck and Tavakolian (1982) provide evidence which implies that the representation of animate entities take up more working memory capacity than inanimate things. Given that this is an assignment task which necessitates the maintenance of unresolved information there may be enough of a memory load to cause this difference between animate and inanimate non-subjects to have an effect. If it is the case that animate non-subjects are causing the processor considerable difficulty then the most basic information for a heuristic to operate on would be trace strength. Gernsbacher, Hargreaves and Beeman (1989) make a case for a recency effect which is overtaken by a primacy effect, after about 1500ms. The subjects in this experiment may have been relying on recency to help them resolve subject pronouns which would cause a bias towards non-subject antecedents. However by the time non-subject pronouns were reached primacy may have become the dominant information in which case non-subject pronouns would expected to be resolved to subject antecedents. This last prediction is only weakly supported by the data in the Animate Static condition and confounded by the results from the Animate Dynamic condition. If simple low level memory functions are being relied upon because of the lack of structure then it may be the case that performance will be chance when non-subject pronouns are encountered. Clearly there is a complicated interaction with Dynamic and Static spatial descriptions. If animate non-subjects are indeed causing the processor to misfunction then it is hardly surprising that the results are difficult to account for.

3.7 Experiment VII

This experiment is closely related to the previous one (Experiment VI) in that it uses similar materials to those used in Experiment III and neither of the noun phrases in the first clause are repeated in the second clause. However, the pronouns are made unambiguous by using number and gender so in that sense it differs from Experiment VI. In the sentences which use inanimate non-subjects one of the potential antecedents is singular and the other plural—the number of the pronoun is then enough to identify the antecedent. In animate sentences the stereotypical gender of the two antecedents is different and therefore the gender of the pronoun can be used to disambiguate the reading. Therefore this experiment is very similar to Experiment III but the point of resolution of the non-subject pronouns has been postponed until the pronoun is encountered.

3.7.1 Subjects

The subjects were 48 volunteer students from Durham University.

3.7.2 Design and Materials

The design was similar to Experiment III in that there were the same four design factors: Animacy (animate *vs.* inanimate), Text (Dynamic *vs.* Static description), Pronoun (subject *vs.* non-subject) and Antecedent (subject *vs.* non-subject). Forty eight sentence sets were constructed and each set contained 16 sentences corresponding to all combinations of the four design factors applied to a prototype sentence. As an example, a reduced sentence set appears in Table 3.7 (examples showing the effects of the Pronoun and Antecedent variables have been left out). The prototype sentences were generated from a set of 48 triples of animate nouns

Animacy	Text	Example sentence
Animate	Narrative	The schoolmistress was put next to the boy and she was pushed close to the prefect.
Animate	Spatial	The schoolmistress was next to the boy and she was close to the prefect.
Inanimate	Narrative	The meter was put next to the monitors and it was pushed close to the dial.
Inanimate	Spatial	The meter was next to the monitors and it was close to the dial.

Table 3.7: Example sentence set for Experiment VII. All the examples contain an unambiguous pronoun in the subject position of the second clause which refers to the antecedent in the subject position of the first clause. The full set also contained sentences with object antecedent and object pronouns.

and a set of 48 triples of inanimate nouns and 48 verb phrase frames. A verb phrase frame comprised a pair of verbs and appropriate prepositional phrases. Each prototype sentence was made up from three animate nouns, three inanimate nouns and one verb phrase frame. No triple of nouns or verb phrase frame was repeated within a sentence prototype or across the set of sentence prototypes. Within each prototype sentence the first two animate nouns were different genders. Table B.4 shows the set of triples. The verb phrase frames are the same as those used in Experiment III (see Table B.2).

All prototype sentences took the form of *The* N_1 *was* V_1 PP_1 *the* N_2 *and [*N_3 *or PRO] was* V_2 PP_2 *the [*N_3 *or PRO]*. The set of 16 sentences was then generated by taking combinations of elements from this sentence prototype and making appropriate substitutions which were determined by the levels of the four factors. Therefore, the Static sentences were generated by leaving out the Vs, the Animacy of the sentences was manipulated by choosing which pair of nouns to substitute. In order to allow the pronoun to be unambiguous and to pick out one of the potential antecedents, the number of the two nouns was always different for inanimate nouns and the gender was always different for animate nouns. Consequently, the antecedent of the pronoun was manipulated by choosing the gender or number of the pronoun. Of course, the Pronoun variable determined which N position in the second clause was replaced by a pronoun.

The design of the questions and the assignment of sentence sets to materials was the same as in Experiment III.

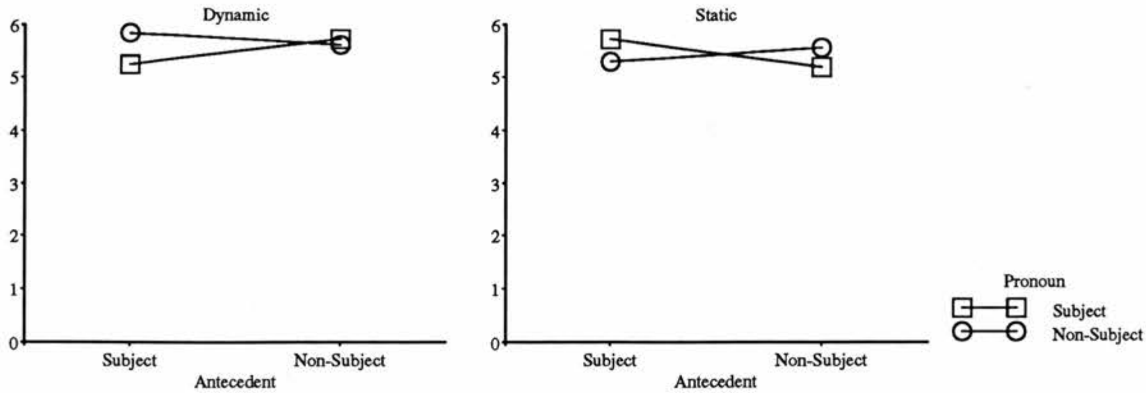

Figure 3.12: Reading rates (in words per second) by Text (Dynamic and Static), Pronoun (Subject and Non-subject) and Antecedent (Subject and Non-subject) for Experiment VII.

3.7.3 Procedure

The procedure was the same as in Experiment III.

3.7.4 Results

Outliers were treated in a similar way to those in Experiment III. Out of 2304 data, 30 (1.3%) data fell below 350ms (mean number per subject=0.625,SD=0.195) and 5 (0.2%) above 10000ms (mean number per subject=0.104,SD=0.472).

Following the procedure in Experiment III all the reading times were converted into rates to remove any artifacts due to the number of words in the comparison between Static and Dynamic sentences. Analysis of the question responses showed that 26% of the answers were incorrect and there did not appear to be any effects of response bias (24% of the questions which required a positive response got a negative response and 28% of the questions which required a negative response go a positive responses). Table B.10 shows the number of correct and incorrect responses by Pronoun position and Antecedent position. Table B.11 shows the number of correct and incorrect responses to questions by all four factors.

Figure 3.12 shows the mean reading rates by Text, Antecedent Position and Pronoun Position. Analysis of variance on the data set with Animacy, Text, Pronoun position and Antecedent position as fixed-factors showed that there were no reliable effects. There is a hint of an interaction between Text, Antecedent position and Pronoun position ($F_1(1,47) = 4.56, p < 0.04, F_2(1,47) = 3.33, p < 0.08$). The effect, which is significant for the subjects analysis seems to show that the Text variable has an effect on the interaction between Antecedent and Pronoun. In Dynamic texts pronouns are quicker to read when they refer to an antecedent in the other grammatical position. However in Static texts, pronouns are quicker to read when they refer to antecedents in the same grammatical position. Further analysis which examined simple main effects and simple interactions revealed that none of the factors, Text, Pronoun or Antecedent had a significant effect at any of the combinations of the other factors. However, there was a hint the interaction between Antecedent and Pronoun is approaching significance ($F_1(1,47) = 3.46, p < 0.07, F_2(1,47) = 2.80, p < 0.2$) at the Static level and not

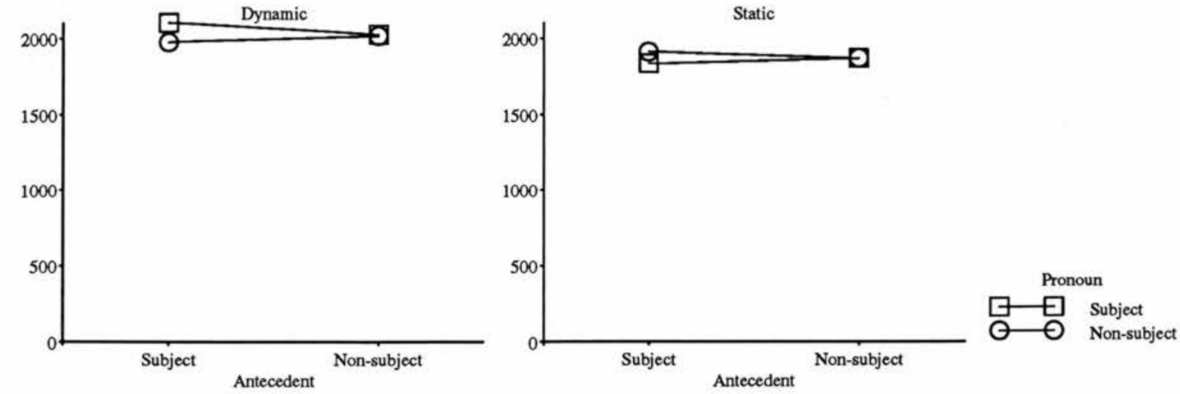

Figure 3.13: Reading times (in milliseconds) by Text (Dynamic and Static), Pronoun (Subject and Non-subject) and Antecedent (Subject and Non-subject) for Experiment VII.

at the Dynamic level ($F_1(1,47) = 2.11, p < 0.2, F_2(1,47) = 1.09, p < 0.4$)

Figure 3.13 shows the mean reading times by Text, Antecedent Position and Pronoun Position. The pattern clearly replicates the pattern of data in Table 3.12. A similar analysis was carried out on the raw reading time data with identical fixed-factors. The analysis showed that there was a main effect of Text ($minF'(1,74) = 5.38, p < 0.05$) with the target clause of Dynamic texts taking on average 2022 ms (SD=1150) to read compared to the target clause of Static texts at 1868 ms (SD=1170). There were no other significant effects either for a subjects or a materials analysis.

3.7.5 Discussion

The reading time analysis confirms that the subjects were reading the texts because there were more words in the Dynamic sentences than in the Static sentences so it was expected that they would take longer to read. Otherwise the pattern of reading time data followed the reading rate data which shows that the transformation was a valid one.

There were no significant results in any of the reading rate analyses of the data but in spite of that the data are still interesting because they formed a clear pattern and the confirmatory statistics did approach significance. The analysis seems to support the view that there is a parallel function effect in Static sentences but not in Dynamic sentences—or rather, there is an interaction between Pronoun and Antecedent position in the predicted way for Static sentences but there is no interaction between Pronoun and Antecedent position for Dynamic texts.

The parallel function effect for Static texts is clearly explicable and replicates the result found in Experiment III for subject pronouns and extends the effect to non-subject pronouns. In a reading time task, if the materials are designed so that the pronouns remain ambiguous up until at least the pronoun is encountered then it seems as if a parallel function is observed for single-sentence Static descriptions.

However, this result does not apply for Dynamic sentences. If anything there is a reverse parallel function effect in which it is faster to assign an antecedent to a pronoun in the contrasting grammatical position. Obviously the difference between the two types of condition relates to the verb in Dynamic sentences. The presence of this verb causes the sentence to become passive and for an unknown agent to be introduced. In the Static sentences the verb is active and there is no unknown agent. It could be the case that the Dynamic sentences simply cause the processor to fail or to adopt some spurious strategy which produces random results. However, there may be an explanation for the results which might indicate that they are not random and that they may reveal some aspect of the way in which these sentences are processed.

Consider Huttenlocher's explanation of why end-anchoring makes three term series problems easier which relates to De Soto, London and Handel's (1965) imagistic theory of series problems. She suggests (Huttenlocher, 1968) that it is not the fact that the first mentioned item of a premise is an end-anchor but that it is the grammatical (deep subject) of the sentence which is crucial. She and others (Huttenlocher & Strauss, 1968; Huttenlocher, Eisenberg & Strauss, 1968) found that in sentences like "The red truck is pulled by the green truck", if the truck which had to be moved was the grammatical non-subject rather than the grammatical subject then children found the situation harder to act out. In order to make their explanation apply to adults she supposes that adults manipulate image-like arrays the way children manipulate real arrays.

Suppose that the subjects in this experiment were comprehending the materials by forming some kind of image. In the Dynamic sentences the verbs describe some kind of movement which is always of the subject being moved in relation to the non-subject. If the real-world object to be moved is the surface subject and deep object then the sentence should be harder to understand. This is the case in all the target sentences of the four conditions of Pronoun by Antecedent (*e.g.*, *she* was pushed close to the prefect where *she* is a surface subject and a deep object). However, if the real-world object to be moved (the antecedent of the pronoun) has already been a surface subject and a deep object (*e.g.*, The schoolmistress was put next to the boy and she was pushed close to the prefect) then that may be harder than if the object has not been mentioned before (*e.g.*, The schoolmistress was put next to the boy and the prefect was pushed close to her) or it has some other status like indirect object (*e.g.*, The schoolmistress was put next to the boy and he was pushed next to the prefect). This case does arise when there is a pronoun in the subject position which refers to an antecedent in the subject position (direct object) of the first clause and if images are being manipulated which are sensitive to deep and surface grammatical position then it (the pronoun) should be hard to process. This case is represented in the data and does show the lowest reading rate. The other three cases are all very similar in value and show that moving an indirect object or a third entity has little effect.

Although Huttenlocher's explanation and data do not relate directly to the data from this experiment it is evidence that a conflict between deep and surface grammatical position can cause processing difficulties. Subjects may also have processing difficulties if they try to "imagine" the agent which moves the described objects about. Essentially the difference that appears in the data between levels of the Text variable may be explained in terms of normal processing and abnormal processing rather than two forms of normal processing.

Pronoun	Example passage
	Conjoined antecedents
	John and Sammy were playing in the garden.
	Ellen watched their game with interest.
Non-Subject	John pushed Sammy and Ellen kicked him.
Subject	John pushed Sammy and he kicked Ellen.
	Antecedents in subject predicate form
	John was playing with Sammy in the garden.
	Ellen watched their game with interest.
Non-Subject	John pushed Sammy and Ellen kicked him.
Subject	John pushed Sammy and he kicked Ellen.
	Antecedents in separate senteces
	John and Ellen were playing in the garden.
	Sammy watched their game with interest.
Non-Subject	John pushed Sammy and Ellen kicked him.
Subject	John pushed Sammy and he kicked Ellen.

Table 3.8: Two versions of the same target sentence wtih three different context sentence pairs for Experiment VIII.

3.8 Experiment VIII

The aim of this experiment is to improve on the design of Experiment V and investigate the same phenomena. In order to collect more data Text was made a between-subjects variable which would also remove any strategic difficulties. A baseline condition was added to the three levels of the Text factor so that it would be easier to assess what the effects of the context sentences in the other three conditions were. Subject pronouns were also included to check for parallel function in comparison with object pronouns.

3.8.1 Subjects

The subjects were 192 volunteer students from Durham University.

3.8.2 Materials and Design

Each subject read 48 passages consisting of two context sentences followed by a target sentence. The design factors were Text type (conjoined antecedents, subject-predicate antecedents, separate antecedents and target sentence only), Participant order (same order *vs.* different order), Pronoun position in target (subject *vs.* non-subject). Half the targets contained 3 or 4 words after the *and* and half the targets contained 5 to 7 words after the *and* making a between materials variable of length with two levels. Table 3.8 shows a reduced text set illustrating three of the levels of the text style variable: conjoined antecedents, subject predicate antecedents and separate antecedents.

The same 48 text sets which were used to make up the materials in Experiment V (see Table B.7) were used in this experiment. Each text set was used to make 16 versions of itself (four levels of the Text variable, two levels of the Order variable and two levels of the Pronoun variable). There were sixteen material sets which were made up of four groups of four materials sets. Each group of material sets contained texts which were all at one level of the Text variable. A group of four material sets was formed by assigning one of the four versions of a text at a particular level of the Text variable to each material set. Therefore across the four material sets of a group each text occurred in all of its four versions. However, within each material set, each text occurred in one of its four versions.

Each text was followed by a question which was designed the same way as the questions in Experiment V.

3.8.3 Procedure

The procedure was identical to the procedure in Experiment V.

3.8.4 Results

Outliers were treated in a similar way to those in Experiment III.

Comparison over three levels of Text (Conjoined, Subject-predicate, Separate)

Assignment data The assignment data was prepared as before by subtracting the number of non-subject assignments from the number of subjects assignments by condition by subject. Analysis of variance (Pronoun, Order, Length and Text) showed that there was one reliable effect of Pronoun where the subject assignment bias was greater for subject pronouns ($minF'(1,163) = 27.3, p < 0.01$, subject=2.3, SD=2.76, non-subject=0.8, SD=2.61). Two interactions approached significance: Order by Text, $F_1(2,141) = 3.95, p < 0.03$,$F_2(2,69) = 4.33, p < 0.02$ and Pronoun by Text $F_1(2,141) = 3.09, p < 0.05$,$F_2(2,69) = 5.04, p < 0.01$. Figure 3.14 and Figure 3.15 show the means for the two interactions respectively.

The interaction between Order and Text appears to be driven by the effect of Order in the Separate condition. Analysis of simple effects shows that there is no effect of Order in the Conjunct and Subject-Predicate texts and a marginally significant effect of Order in Separate texts ($F_1(1,141) = 6.86, p < 0.01$, $F_2(1,69) = 5.58, p < 0.03$, $minF'(1,171) = 3.08, p > 0.05$). The interaction between Pronoun and Text seems to be caused by the effect of Text on Non-Subject pronouns rather than Subject pronouns. Simple effects analysis shows that at the Subject level of Pronoun there is no effect of Text and a marginally significant effect at the Non-Subject level ($F_1(2,141) = 5.09, p < 0.008$, $F_2(2,69) = 4.49, p < 0.02$, $minF'(2,177) = 2.39, p > 0.05$).

Reading Times As before reading times over 10000ms and under 350ms were replaced. A maximum likelihood analysis was carried out on the data with Pronoun, Assignment, Text, Order and Length as fixed factors. There were four significant main effects which are described in Table 3.9. There was one significant interaction which was between Pronoun, Assignment,

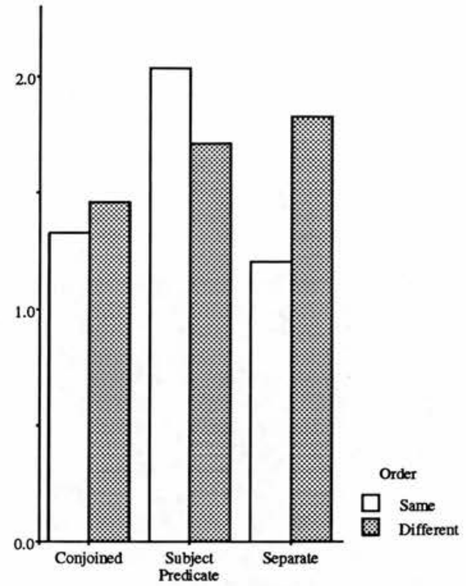

Figure 3.14: Pronoun assignment bias by Order (Same and Different) and Text (Conjoined, Subject-Predicate and Separate) in Experiment VIII where N=192.

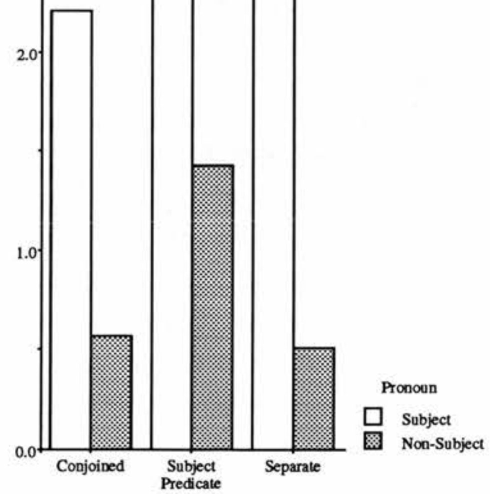

Figure 3.15: Pronoun assignment bias by Pronoun (Subject and Non-subject) and Text (Conjoined, Subject-Predicate and Separate) in Experiment VIII where N=192.

Factor	F_1	$p<$	F_2	$p<$	Level	N	Mean	SD
Pronoun	22.59	0.0001	11.76	0.0007	Subject	3456	2450	1615
					Non-subject	3456	2346	1461
Assignment	22.31	0.0001	5.05	0.03	Subject	4372	2360	1476
					Non-subject	2540	2462	1644
Text	11.66	0.0002	198.08	0.0001	Conjoined	2304	2132	1295
					Subject-Predicate	2304	2125	1382
					Separate	2304	2936	1761
Length	131.13	0.0001	9.38	0.003	Short	3456	2228	1462
					Long	3456	2567	1598

Table 3.9: Significant main effects and means for levels in Experiment VIII over Conjoined, Subject-Predicate and Separate levels of Text. The degrees of freedom are 1 and 6864 except for Text where they are 2 and 6864.

Order and Length ($F_1(1,6864) = 5.77, p < 0.02$, $F_2(1,6864) = 9.38, p < 0.003$), the means of which are presented in Figure 3.16.

Comparison over all four levels of text (Conjoined, Subject-predicate, Separate and Single

Assignment Data Because the last condition of text has only one sentence the order factor (levels Same and Different) clearly does not apply so to make a comparison over text the Order variable has been collapsed over. As before the assignment data was prepared for analysis of variance by subtracting the number of non-subject assignments from the number of subject assignments by condition for each subject. An analysis of variance was then done on the data using three fixed factors: Text (Conjoined, Subject-Predicate, Separate, Single), Length (short, long), Pronoun (subject, non-subject). The analysis showed that there were two reliable effects. Subject pronouns received more assignments to the subject antecedent than to the non-subject antecedent ($minF'(1,83) = 32.32, p < 0.01$, subject=4.7, SD=4.5 and non-subject=1.2, SD=1.2). There was also an interaction between Pronoun and Text, $minF'(3,203) = 2.89, p < 0.05$. Figure 3.17 shows the means.

Analysis of simple effects shows that Text has no effect at the subject level of Pronoun (F_1 and $F_2 < 1$) but does have a reliable effect at the non-subject level of Pronoun ($minF'(3,187) = 4.36, p < 0.01$). This shows that subject pronouns are largely unaffected by the preceding sentences whereas the subject assignment bias for non-subject pronouns is affected. The bias is largest when the two potential antecedents are introduced in a subject predicate form which is attenuated if the antecedents or conjoined subjects or subjects introduced in separate sentences. The smallest bias which is slightly negative (an non-subject assignment bias) is for target sentences presented without any context.

Figure 3.18 shows the interaction between Text and Pronoun broken down by structure in the target sentence of the materials: some had parallel structure so grammatical parallelism could operate and the rest had non-parallel structure.

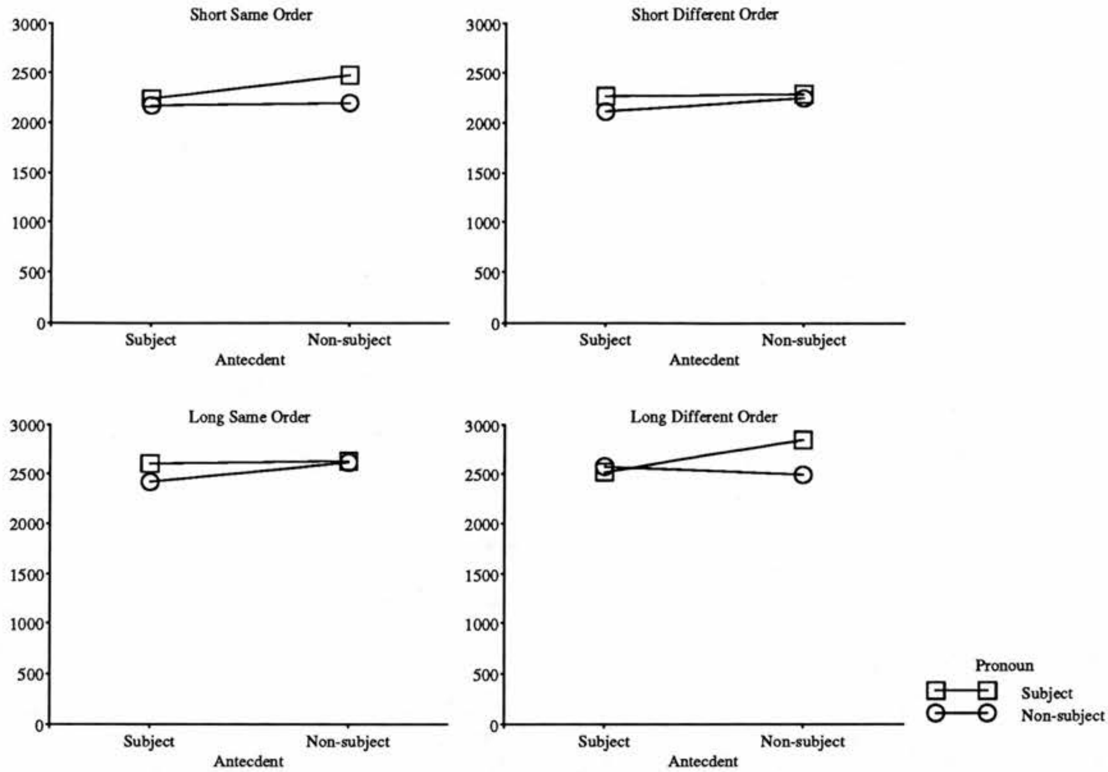

Figure 3.16: Reading times (in milliseconds) by Pronoun (Subject and Non-subject), Assignment (Subject and Non-subject), Order (Same and Different) and Length (Short and Long) from Experiment VIII.

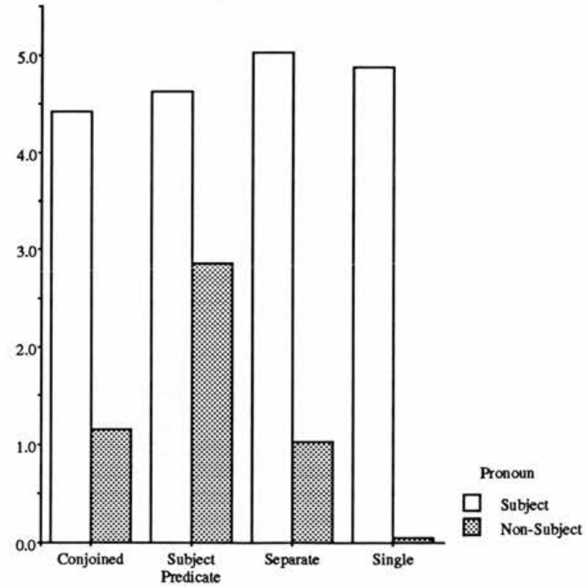

Figure 3.17: Assignment biases by Text (Conjoined, Subject-Predicate, Separate and Single) and Pronoun (Subject and Non-subject) in Experiment VIII where N = 96).

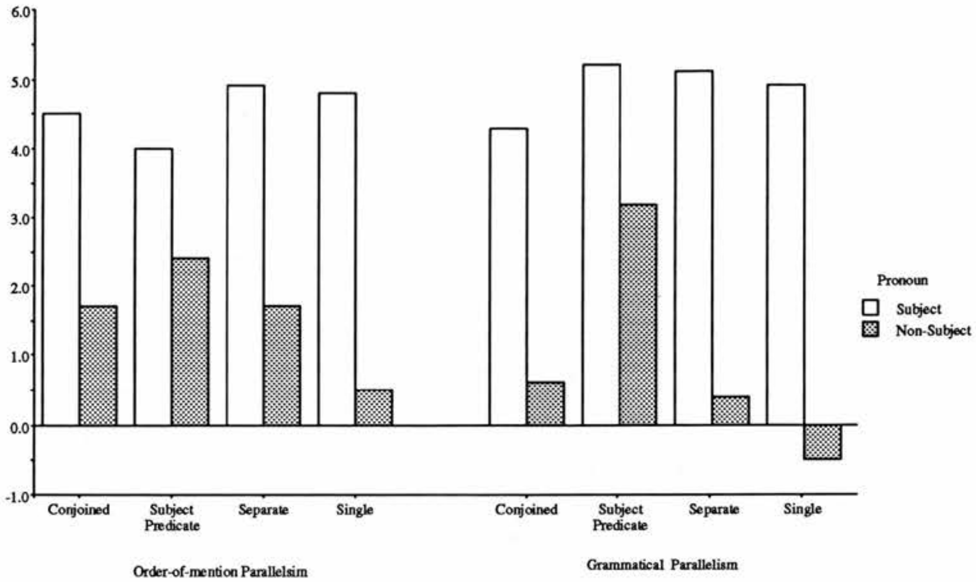

Figure 3.18: Assignment biases by Text (Conjoined, Subject-Predicate, Separate and Single) and Pronoun (Subject and Non-subject) in Experiment VIII for the parallel and non-parallel groups of materials (averaged of subjects to give cell totals of 12. N = 46 for non-parallel and N = 50 for parallel).

Reading Times As before reading times over 10000ms and under 350ms were replaced. A maximum likelihood analysis was carried out on the data with Pronoun, Assignment, Text and Length as fixed factors. There were four significant main effects which are described in Table 3.10. There was also an interaction between Text and Pronoun ($F_1(3, 9183) = 6.76, p < 0.0002$, $F_2(3, 9183) = 5.78, p < 0.0007$) which is described in Figure 3.19.

Factor	F_1	$p<$	F_2	$p<$	Level	N	Mean	SD
Pronoun	10.67	0.002	3.40	0.07	Subject	4608	2402	1614
					Non-subject	4608	2364	1510
Assignment	51.60	0.0001	11.36	0.0008	Subject	5756	2334	1480
					Non-subject	3460	2466	1689
Text	6.99	0.0002	126.91	0.0001	Conjoined	2304	2132	1295
					Subject-Predicate	2304	2125	1382
					Separate	2304	2936	1761
					Single	2304	2340	1628
Length	6.76	0.0002	5.78	0.0007	Short	4608	2206	1484
					Long	4608	2561	1620

Table 3.10: Significant main effects and means for levels in Experiment VIII over Conjoined, Subject-Predicate, Separate and Single levels of Text. The degrees of freedom are 1 and 9183 except for Text where they are 3 and 9183.

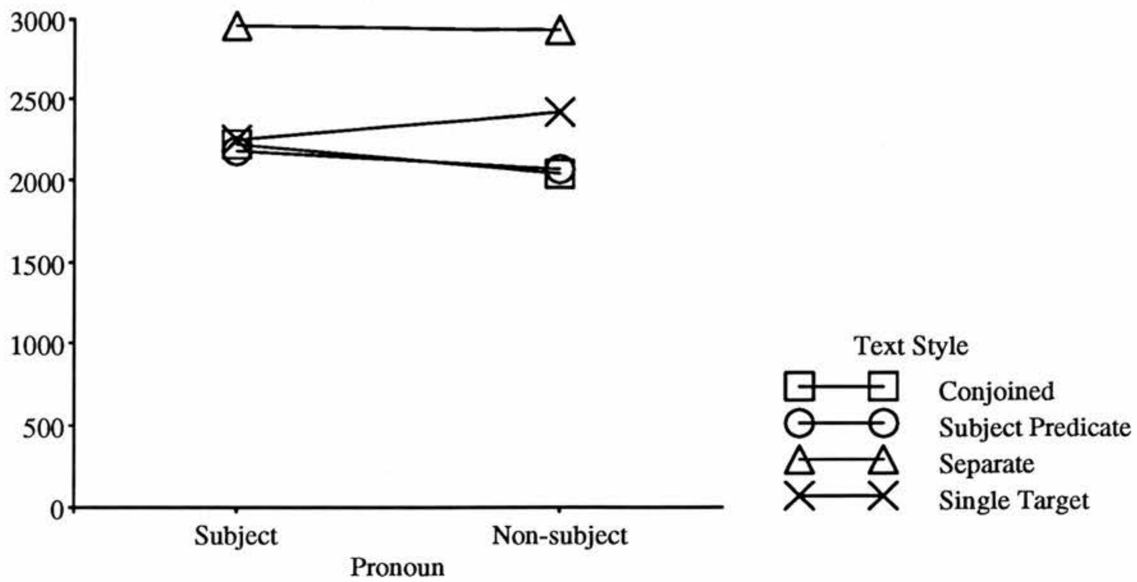

Figure 3.19: Mean reading times (in milliseconds) by Text (Conjoined, Subject-predicate, Separate and Single) and Pronoun (Subject and Non-subject) in Experiment VIII where N=1152.

3.8.5 Discussion

Assignment data

Both analyses show a strong subject assignment bias which agrees with Crawley, Stevenson and Kleinman's (1990) results. The first analysis over three levels of the Text variable shows an interaction between Order and Text. The interaction demonstrates that Order has no effect when the antecedents are both presented for the first time in the first sentence (the Conjoined and Subject-Predicate texts). However, Order does have an effect in Separate texts where the antecedents are introduced in consecutive sentences which is curious at first sight: the size of the subject assignment bias is attenuated when the Order is Same compared to the Different Order.

The analysis of the three Text conditions showed an interaction between Order and Text which was caused by an effect of Order in the Separate condition. This effect may be explained by considering topic effects. Many investigators (*e.g.*, Caramazza & Gupta, 1979) have claimed that the surface subject of a sentence is topicalised and that a new subject in a new sentence introduces a new topic (*e.g.*, Karmiloff-Smith, 1985). Therefore, when the Order is the Same in a Separate text, the second introduced individual (which will be the object antecedent in the target) becomes the most recent topic. This means that if a topic assignment strategy is used to resolve the pronoun, as well as a subject assignment strategy, then the two strategies disagree (grammatical parallelism has no impact because there is no interaction with Pronoun). However, when the Order is Different the antecedent mentioned in the second sentence which will be topicalised, is the antecedent which will be in the subject position of the target. Therefore the two strategies (topic assignment and subject assignment) will agree. This difference between the two assignment strategies accounts for the effect of Order at the Separate level of Text.

Terry was going to meet Hector at the cricket match.
Their friend Brenda came along to watch as well.
Terry took Hector to the pavilion and he waved to Brenda.

Table 3.11: An example of a text with non parallel clauses in the target sentence.

This argument does not apply to the other two Text conditions because in those two conditions the individual introduced in the second context sentence is a third individual who is never a potential antecedent of the pronoun. Thus, the effects of the most recent topic that were observed in the Separate condition, are not apparent here and in addition the effect of the initial topic are presumably attenuated.

There is an interaction between Pronoun and Text in both the analyses which is essentially the same because the pattern of results in the analysis over three levels of Text is clearly repeated in the analysis over all the Text levels. The assignment bias for Subject pronouns remains constant over all levels of Text but Non-subject pronouns are sensitive to the Text structure. The subject assignment bias is strongest for Subject-Predicate texts and virtually negligible for Single texts and in between for Conjoined and Separate so there appears to be an ordering of parallelism effects from Subject-Predicate, through Conjoined and Separate to Single. Several observations are apparent. First, context sentences before the target sentence increase the subject assignment bias (Conjoined, Subject-Predicate and Separate levels of Text compared to Single). Second, when the antecedents are assigned different grammatical functions in the first context sentence the subject assignment bias is larger than in the conditions when they are assigned the same grammatical function, regardless of their positions in the context. Perhaps, the subject antecedents introduced in the Subject-Predicate texts have their subjecthood emphasized by the contrast expressed in the first sentence and hence the subject assignment bias is larger in those texts.

On inspection of the materials it is apparent that some of the target sentences have parallel constituent structure whereas others do not. Table 3.11 shows an example text which has a target that does not have parallel constituent structure. The tabulation of the interaction between Text and Pronoun for both groups (parallel and non-parallel targets) shows that this difference does have an effect (see Figure 3.18). For Non-subject pronouns, in Conjoined, Separate and Single texts the subject assignment bias is less in the parallel group than in the non-parallel group which is exactly what would be expected if grammatical parallelism were unable to operate in the non-parallel targets. It is also evident that the residual effect in the non-parallel texts must be be due to parallelism operating over order-of-mention or a weak sense of grammatical function (subject *vs.* non-subject).

In spite of the clarifying nature of the tabulation a new puzzle is apparent. The difference between the two groups of targets is reversed for Subject-Predicate texts: the subject assignment bias is greater in the parallel targets compared to the size of the bias in the non-parallel targets. Unfortunately there is no clear explanation of this result and indeed it may be a freak result.

In summary, the Order by Text interaction lends some support to an order of mention strategy operating over texts but the explanation is probably a recency one because the effect does not interact with Pronoun. The interaction between Pronoun and Text when divided by parallel and non-parallel targets provides evidence for inter-sentential parallelism and intra-sentential parallelism. Finally, all the data shows a strong subject/first-mention effect and that subject pronouns show very few effects other than this strong subject assignment strategy. This could be a very strong form of parallel function or it could indicate something about sentence processing. Obviously subject pronouns arrive early in a sentence and an interpretation of the sentence has not yet been formed so the best strategy to adopt may be a subject/first-mention one because the rest of the sentence's structure has yet to be determined. However, the case for Non-subject pronouns is different because they are processed much later and there is much more scope for other sources of information to interact to form an interpretation because most of the sentence has been interpreted.

Reading Times

In the analysis over the three levels of Text there is an interaction between Length, Order, Pronoun and Assignment. It is immediately clear that the pattern of the Pronoun by Assignment interaction is very similar in the Short Same and Long Different condition and likewise in the Short Different and Long Same condition. The influence of the Order and Pronoun factors shows that grammatical parallelism is operating alongside and sometimes in competition with an order-of-mention parallelism. The Length factor is perhaps a little more mysterious but may be related to the distinction described above, between parallel and non-parallel targets. For the short materials nine out of 24 have parallel targets and in the long materials 13 out of 24 have parallel targets. This difference may cause the average effects of parallel function to differ in the two length conditions. If it is assumed that parallel function is used rarely in the Short condition then the explanation for the short half of the analysis is relatively straightforward. In the Same condition a subject assignment strategy and a parallel order strategy both agree giving a speed advantage to subject pronouns. This situation is reversed for the Non-subject pronouns canceling out and showing no time difference for the two assignments. In the Different condition the situation is reversed for the two levels of Pronoun so that a symmetric pattern is observed. Unfortunately the pattern of results in the Long condition is less clear although is does appear to be a reflection of the pattern in the Short condition.

The main effects in the first and second analysis show the same pattern. The main effect of Pronoun shows that non-subject pronouns are faster to process than subject pronouns which may be related to the point made above about the construction of coherent interpretation. Assignments to subject antecedents take less time than assignments to non-subject antecedents which is expected given the large subject assignment bias in the assignment data (main effect of Assignment). The pattern of times over Text (main effect of Text) reflect some of the similarities in the assignment analysis which showed Conjoined and Subject-Predicate texts were similar (in the Order by Text interaction) whereas Separate had curious recency effects which may cause the delay and Single texts provide a benchmark measure which demonstrates that the Conjoined and Subject-Predicate texts have a facilitatory effect whereas Separate texts hinder processing.

The analysis over all four levels of Text showed only a single interaction between Text and

Assignment. The pattern of results shows that Non-subject assignments are quicker in all cases except the Single condition where the expected result is observed. This ties in with the point about forming a coherent whole because the difficulty in satisfying the multiple constraints of a text are clearly minimised in the Single condition so Non-subject pronouns will no longer have an advantage by appearing late in a sentence.

Comparison of Assignment and Assignment-Times Analysis

Several points emerge in the comparison. It appears that parallelism over order-of-mention appears only in the assignment times because there is an interaction between Pronoun and Order but not in the assignment analysis. Text level parallelism effects emerge in the assignment analysis but not in the times analysis. Intra-sentential grammatical parallelism appears in both analyses as does the effect of forming a coherent interpretation. This effect is particularly interesting for explaining the much neglected difference between subject and non-subject pronouns. The effect is intuitively appealing given the left-to-right incremental nature of interpretation and the evidence discussed here supports it to some extent.

3.9 General discussion of Experiments VI, VII and VIII

3.9.1 Comparison of Experiments VI and VII

Consider first Experiments VI and VII. These two experiment were designed to overcome the problems of Experiment III. In Experiment III it was clear that the crucial non-subject pronouns were not being treated in the same way as the subject pronouns: they were effectively being resolved after the first noun phrase of the second clause which meant that the RTs revealed nothing about subjects' processing abilities. To overcome this problem a third noun phrase was introduced in the second clause. This meant that in Experiment VI, where the two potential antecedents were the same gender and the same number, that the pronoun was ambiguous. Of course, this equalized the way in which subject and non-subject pronouns were treated in that they were both equally ambiguous. The data clearly supports this conclusion because both non-subject and subject pronouns show an effect of antecedent position on assignment.

The materials used in Experiment VII were very similar to those in Experiment VI except that the pronouns were made unambiguous. This was done by making the gender or the number of the two potential antecedents different so that the gender or number of the pronoun was sufficient to pick out its antecedent. Although the pronouns were unambiguous, as in Experiment III, these materials equalized the point of pronoun resolution by making the pronouns unambiguous when they were encountered. The data again show that this variation of the materials from Experiment III was successful in revealing effects of antecedent position on non-subject pronoun resolution. Presumably the RTs revealed how easy or hard it was for the processor to make the appropriate assignments.

Both Experiments (VI and VII) showed how the introduction of the third noun phrase in the second clause of the sentences overcame the differences between subject and non-subject pronouns found in Experiment III. Furthermore, the addition of the third noun phrase made the sentences easier to understand because the question answering accuracy increased from Experiment III to Experiment VII even though the verbs and original noun phrases were retained.

Although these two experiments have revealed interesting information about the relationship between subject and non-subject pronouns, their results, taken together, are curious. In Experiment VI there is an effect of Animacy on pronoun resolution and no effect of Text but in Experiment VII there is an effect of Text (although only approaching significance) and not an effect of Animacy. Bearing in mind that the materials in the two experiments are very similar this difference is a curiosity.

Of course, the explanation of the differences must lie in the differences between the two tasks: one is an assignment task while the other is a reading task. An assignment task does not provide any information about the processing of a sentence while it is being read and a reading time task can only provide information about the difficulty of comprehending a sentence. Therefore the two experiments imply that Text may effect the ease which the processor has in making an assignment while Animacy has no effect, but if the system is given the extra load of maintaining unresolved information, as in the assignment task, then Animacy can affect the final outcome. Unresolved information may impose an extra load in the sense that the processor may actually maintain two possible readings in the "hope" of encountering disambiguating information. Of course, Text style may have an effect during the processing of sentences in the assignment task but it does not affect which assignment is made. The reason Animacy does not have any effect in the reading time task is because there is no need to maintain unresolved information because the pronouns are unambiguous.

In these two experiments where there is an effect of Text or Animacy, one level produces the predicted behaviour or parallel function; the Inanimate level of Animacy in Experiment VI and the Static level of Text in Experiment VII. Both these levels seem to be the unmarked level of their respective factors, because in a sense, they are more normal than their opposing level (the Animate level of Animacy and the Dynamic level of Text). The materials describe spatial layouts and they may be dynamic or static which is perfectly normal state of affairs. However, it does seem unusual to describe a dynamic spatial layout without mentioning the agent which moves the objects whereas, it is more normal to simply state a spatial arrangement as in Static sentences and not mention any possible agent and so the Dynamic level is marked. In the case of animate non-subjects it is unusual to simply describe their positions, usually animate objects do things whereas inanimate objects are not generally agents so the Animate level is marked.

This difference between Dynamic and Static, and Animate and Inanimate may help to interpret the results in the "marked" conditions; Dynamic and Animate. In these conditions the results are hard to interpret and there is a question whether they are spurious. It could be that the processor is behaving in an unpredictable way which happens to have produced these results which are random as the confirmatory statistics imply. Of course, they could be revealing, but given the oddness of the materials, the difficulty of interpretation and the lack of statistical support it is probably safe to treat with caution the results from the Dynamic condition of Experiment VII and from the animacy condition of Experiment VI.

The results from the Static condition in Experiment VII and from the Inanimate condition in Experiment VI are not quite in agreement. For subject pronouns they both show an advantage for the subject antecedent: in Experiment VII the clause is read faster if the subject pronoun refers to the subject antecedent and in Experiment VI a subject pronoun is more likely to be assigned to the subject antecedent. These results are consistent with a subject assignment strategy or a parallel function strategy which is consistent with previous experimental findings (*e.g.*, Grober, Beardsley & Caramazza, 1978; Fredriksen, 1981). The results for the non-subject pronouns are less straightforward. In Experiment VII non-subject pronouns are processed faster when they refer to non-subject antecedents than when they refer to subject antecedents and this "parallel" advantage is similar in size to the subject pronouns advantage. However in Experiment VI non-subject pronouns show a very small bias to non-subject assignments which is much smaller than the bias to subject assignments for subject pronouns: in fact, there may be no bias for non-subject pronouns.

In an attempt to explain these differences the nature of the two tasks should be taken into consideration. The reading time results may reflect something about the nature of the representations being manipulated by the processor in that parallel structures take less time to map onto each other but not say anything about what heuristics might be used. The assignment data directly reflects peoples biases and it is clear that there are two explanations for these observed biases. It may be that there is a subject assignment strategy operating alongside a parallel function strategy: they agree on subject pronouns and disagree for non-subject pronouns. Taking the two experiments together it does seem more likely that there are two heuristics operating together because if there is a representational advantage in parallel structure then one would expect language to exploit it and therefore there to be a language regularity for a heuristic to exploit.

3.9.2 Comparison of Experiment VIII with Experiment V

Experiment V used only non-subject pronouns and manipulated Text as a within-subjects factor. Because it used only non-subject pronouns there was no way of revealing any parallelism effects and Text may have caused strategic confusions in subjects. The number of subjects was also quite small which meant that analysis was difficult because of empty cells. Experiment VIII remedied these problems and revealed parallelism effects by including subject pronouns as well and it made analysis easier because of the much greater number of subjects. It also examined the effects the context sentences had by including a single sentence condition which showed that context sentences increased the subject assignment bias.

3.9.3 Comparison of Experiment VIII with Experiment VI and Experiment VII

Contrasts

The results of Experiment VIII apparently contrast with those of Experiment VI as mentioned in Section 3.6.5. The results from Experiment VI show that Animate antecedents cause unusual pronoun resolution choices. In Experiment VIII all the potential antecedents are Animate and yet the reliable effects are clearly interpretable. Presumably the difference lies in the fact that the materials in Experiment VIII are more natural and provide a clearer context for interpreting the pronouns against. In Experiment VI there is little narrative structure to help in structuring the information which may cause an increase in memory load.

Similarities in assignments

There are also similarities between the sets of results. In the assignment data from Experiment VI and Experiment VIII there is a clear subject assignment strategy operating and the effects of the design variables operate in competition with it. Sometimes the strategies agree and at other times they do not but the subject assignment strategy is the most influential. In all but one of the conditions in Experiment VIII (Non-Subject level of Pronoun in Single level of Text) the pronoun is most often assigned to the subject antecedent and in Experiment VI when the parallel function strategy competes as in the Inanimate Dynamic and Static conditions the subject assignment strategy is still operating. Perhaps because Experiment VIII uses texts of three sentences the notion of Topic is more significant and therefore accounts for the subject assignments strategy's greater influence.

Both sets of assignment data suggest that a parallel grammatical function strategy is operating because the grammatical function of the pronoun does have an effect in both experiments. In Experiment VI non-subject pronouns are treated differently from subject pronouns and in the Inanimate condition show a slight non-subject assignment bias and in Experiment VIII non-subject pronouns decrease the subject assignment bias in all conditions.

Similarities in reading/assignment times

The comparison between assignment and reading times is not straightforward. There is no main effect of Pronoun in Experiment VII but there is in Experiment VIII which can be explained by the essential difference between the two experiments: in Experiment VII there is no need to "pick" an antecedent, the reading time reflects the "ease" with which an assignment is made whereas the times in Experiment VIII reflect the time to pick an antecedent and the ease of making the assignment and then maintaining that information until the end of the target clause. In both experiments non-subject pronouns show an effect of antecedent position in all conditions although the direction of the effect is not constant: in some conditions there is support for a parallel grammatical function and in others the reverse seems to be the case.

3.9.4 Summary

Experiments VI and VII both used materials in which the potential antecedents had different grammatical functions and the order of subject and non-subject in the two clauses were always the same, which meant that grammatical function was confounded with order of mention. The materials were all single sentences which meant that the effects were all intra-sentential ones.

Experiment VI showed that when the potential antecedents were inanimate there was a grammatical function effect where subject pronouns were assigned to subject antecedents more often, and non-subject pronouns to non-subject antecedents more often. When the potential antecedents were Animate then subject pronouns were assigned to non-subject antecedents more often than subject antecedents and this effect was larger for subject pronouns. The results showed that animacy affects subject pronouns and that non-subject pronouns are largely unaffected.

The results also suggested a more complicated set of effects and implied that the results from the animate sentences were spurious. The results for the inanimate materials were much clearer and suggested that there were two strategies operating: subject assignment/advantage of first mention and grammatical parallelism/order-of-mention parallelism.

None of the results from Experiment VII were confirmed by statistics. However, there was a fairly clear pattern which approached significance. In Static texts there was a parallel function effect where pronouns were faster to process if they referred to antecedents with the same grammatical function (order of mention). The situation was reversed for Dynamic texts but there are reasons for doubting those results, related to the difference between deep and surface grammatical function.

Experiment VIII produced a complicated pattern of results which were disrupted by difficulties with the materials. However, these difficulties were useful in separating inter- and intra-sentential grammatical parallelism. There was a considerable bias to assigning pronouns to the subject antecedent and subject pronouns showed very little influence of the designed variables whereas non-subject pronouns were clearly sensitive. Context sentences influenced the assignments by introducing possible recency effects and strengthened the subject assignment bias. There was very little evidence for order-of-mention effects in the assignment data.

The time to make the assignments in Experiment VIII showed several reliable effects. The main effects were consistent with the assignment results and the effect of context sentences. The interactions in the first analysis over the first three levels of Text is slightly curious. However, the interaction does suggest that order-of-mention effects are operating in the assignment times and the effect of Length further suggest that intra-sentential grammatical parallelism was interacting as well.

3.10 General Discussion

The previous work on parallel function, reviewed above, generally found a subject assignment effect although Cowan (1981) did claim to have found a parallel function effect in pronoun assignment and Frazier, Taft, Roeper, Clifton and Ehrlich (1984) observed a parallel structure effect. Essentially, this work showed that parallelism over structure or at least grammatical function could be exploited in text processing. The work of Gernsbacher (Gernsbacher & Hargreaves, 1988; Gernsbacher, Hargreaves & Beeman, 1989) showed that primacy and recency could both be observed in simple sentences and that the availability of the information varied with time. Therefore taking the two bodies of work together, there is evidence that parallel function over order of mention could be used in text processing because information about order is there (as demonstrated by Gernsbacher's work) and parallel function has been observed in pronoun resolution and other text processing tasks.

The immediate aim of the experiments described here, was to investigate parallel function in pronoun resolution because its status as a genuine heuristic was unclear and to try to separate those effects from order of mention effects. Ultimately the aim was to see if parallelism over order-of-mention across a text could be observed which would provide support for the conclusion drawn at the end of Chapter 2, that parallelism over order-of-mention was a general property of the representational mechanisms used in text processing.

First, consider the evidence for parallel function provided by these experiments. Experiment VI showed a subject assignment bias which was reduced by object pronouns compared to subject pronouns. The effect was most clearly interpretable in Inanimate sentences. This result clearly shows that object pronouns are behaving differently and possibly modifying a subject assignment heuristic. However, parallel function for object pronouns does not outweigh a subject assignment strategy. So the evidence strongly suggests that parallel function is operating in Inanimate sentences. In Experiment VII none of the effects were confirmed.

However, there is a very suggestive parallel function effect in Static sentences. These two results taken together provide strong converging evidence that parallel function is used in pronoun resolution tasks. Furthermore, Animacy has an effect on assignment tasks such that animate nouns appear to cause a working memory overload forcing the processor to use simple ordering information—in this case, recency information. In reading time tasks Animacy has no effect but the kinetics of the sentence do appear to disrupt parallelism. There is also one caveat that in these two experiments that order-of-mention and grammatical function cannot be distinguished. The only way to separate the two is to make two nouns have the same grammatical function which is the manipulation which Cowan (1981) tried or to make two clauses non-parallel with respect to grammatical function but parallel with respect to order-of-mention.

Experiment VIII uses the conjoined noun phrase manipulation but in a three sentence text which avoids the difficulties that Cowan (1981) reported and relates directly to Crawley, Stevenson and Kleinman (1990). Although the pattern of assignments is complicated there is an effect of Order. However, the effect is most easily interpreted as a topic effect and so there is little evidence for a text level order-of-mention parallelism effect although it is not actually disproved. Experiment VIII also used the other method of separating order-of-mention from grammatical parallelism although in a *post hoc* analysis. Roughly half the target sentences were parallel over order-of-mention and half were roughly parallel over order-of-mention and grammatical function. The tabulation of the results showed a residual parallelism effect when grammatical function parallelism could not operate which was interpreted as evidence for intra sentential order-of-mention parallelism.

Finally consider the comparison of referential change, between these simple sentences with pronouns (which will be referred to as pronoun texts) and the texts used in the MIT. Such a comparison illustrates the related nature of the two sorts of materials used in the experiments reported in Chapter 2 and here. When a reader is presented with a simple sentence which contains an ambiguous pronoun the comprehension task demands that the reader satisfies a set of soft constraints which determine the pronoun's interpretation or reference. These constraints may operate over pragmatics, semantics or syntactic features like grammatical function. In the MIT the situation is very different because definite anaphors are used which determine their reference uniquely so no constraint satisfaction problem is encountered.

When reference changes in MIT texts the reader is not faced with the problem of determining the reference of the anaphor but in the pronoun texts that is a problem when the pronoun is encountered—so in a sense reference change in more complicated in the pronoun texts.

There is a second aspect of referential change which varies across the different texts which have been investigated—namely, predictability. In the predictable MIT texts, a reader knows when reference will change and to which individual it will change. In the unpredictable cases the reader cannot know (at least during the majority of a text's sentences) when reference will change or which individual it will change to. The pronoun texts are similar to the predictable texts in that the reader will now that an anaphor will be presented in the second clause of the target sentence. They are also similar to the unpredictable texts in that the reader will not know to which individual the expected anaphor will refer.

Therefore there is an intuitive continuum of predictability of reference change from predictable MIT texts to unpredictable MIT texts. Somewhere between these two end-points lie pronoun texts in which reference change is predictable in one sense and unpredictable in another. It might be useful to re-express this continuum as a scale of information availability. In predictable MIT texts the reader is provided with information about when reference will change, to which individual reference will change. In the unpredictable MIT texts the reader is not provided with any information about when reference will change or to which individual reference will change. In both these cases though the reader is provided with enough information to determine the reference of the anaphor uniquely. However, in pronoun texts the reader is provided with very little information which only helps to determine the reference of the pronoun and does not allow the reader to anticipate the referent or to fully determine where the pronoun will appear.

In conclusion, the experiments reported here have fulfilled the two aims described above. First, parallel function is used in pronoun resolution which demonstrates that parallelism is exploited at some representational level. Second, parallelism over order-of-mention is also used or at least, given the weakest interpretation, order-of-mention effects affect text processing. These experiments therefore constitute evidence that parallelism of structure at many levels is exploited in text processing and may well be a representational primitive. Because parallelism has been found in two radically different tasks (the MIT and pronoun resolution tasks) the issue of parallelism in text is clearly an important one and possibly an omnipresent one.

Chapter 4

Overt rehearsal in the MIT

Summary

The study of unpredictable reference change in the construction of representations of individuals revealed traces of phonological memory. The interpretation of these effects is not straightforward. As a way of further investigating these effects subjects were asked to externalise their rehearsal. The results suggested at least one interpretation for the original effects. This support from a different measure provided strong evidence that indeed phonological short-term memory was being used in the construction of representations if individuals.

4.1 Introduction

4.1.1 The role of articulatory rehearsal in the MIT

In Chapter 2, two regression models were presented for reading time data which showed word length effects of accumulated descriptions in two MITs. The effects were described but their discussion was postponed until now. The effects were clearly related to status and to changes of reference. Chapters 2 and 3 emphasised the importance of sequence and showed how status restored the transparent mapping from sequence to semantics. Given the relationship between articulatory rehearsal and status perhaps rehearsal was being used in the manipulation of sequence. This would be quite consistent with the loop model which does have sequential properties and could be exploited in maintaining orderings over items (Baddeley, Vallar & Wilson, 1987). The reading time model does not directly address this issue. However, it is clear that another store is being used because the semantic properties of the descriptions (match structure) are predictive and there is an issue about the relation between phonological working memory (the articulatory loop) and semantic working memory or short term memory which the negative word length effects may address.

The word length effects did change between the two models but the most reliable and largest effect was a negative effect. As part of the general process of understanding the use of articulatory rehearsal in an MIT it was clearly necessary to develop an interpretation of this negative effect. At first pass a negative coefficient in a regression model might be thought of as representing negative time. This would be the case if the variables represented component processes of a purely sequential system. However, in this model, variables represent structural aspects of the whole reading process and a negative coefficient represents an acceleration of that whole process. The model never actually predicts a negative time because the other positive coefficients in the model outweigh the negative effects.

The negative coefficient in question operates when there is a switch of reference to the primary individual. The time taken to read such a sentence is reduced in proportion to the number of syllables in the non referenced (secondary) individual's description. The more syllables in the secondary individual's description (all other variables being the same) the faster the sentence will be read. Clearly this acceleration represents a competition between two processes. Many of the other variables represent a processing load related to the processing of the primary individual while this negative effect represents an acceleration of these processes by competition from another process related to the secondary individual.

Because the secondary individual's length in syllables is important it is reasonable to assume that the secondary individual's description is being maintained in the articulatory loop (which is sensitive to time-to-say). If this is the case then its representation will decay unless refreshed. Suppose that the description must be processed before it decays and that the articulatory processes are otherwise engaged, perhaps in encoding new visual input, then the available time is constant. The longer the description is (in phonological terms) the more has to be done in processing it before it decays so time may have to be saved from other processes. The saving in time could be achieved in two ways: either the machinery could be accelerated or parts of the usual processing could be missed out. Perhaps the rehearsal of the primary individual is abbreviated whereas the rehearsal of the secondary individual is not because the secondary individual relies on the phonological store whereas the primary individual relies on a more permanent semantic store.

4.1.2 Methodology

In order to try to investigate these possible explanations of the negative word length effect described above, some measure other than reading times needed to be used. The obvious choice was to ask subjects to externalise their rehearsal while performing an MIT and record their protocols. This data also has the benefit of potentially explaining subjects' strategic use of rehearsal because the order of items to be rehearsed can be analysed.

Ericson and Simon (1984) discuss the collection of Thinking-Out-Loud (TOL) data and its interpretation. Although they do not discuss explicitly the collection of overt rehearsal they emphasise the importance of using a model to interpret the data. For the purposes of this investigation the model used was Baddeley's (1986) model.

Subjects were asked to vocalise any rehearsals which they noticed they were making while performing a normal MIT. The meaning of rehearsal was not given explicitly for fear of interfering with the subjects usual behaviour. Fischler, Rundus and Atkinson (1970) had found that restricted overt rehearsal encouraged processing which maintained information in long term memory (LTM) and free overt rehearsal encouraged processing which maintained information in short term memory (STM). It was preferred to encourage processing which maintained information in STM because that would increase a subject's opportunity of vocalising it. Unfortunately, subjects might take advantage of the increase in the maintenance of STM information and change their strategies consequently. The design consisted of an initial condition in which subjects performed an MIT to get accustomed to the task and to be in a position to notice any covert rehearsals that they were making. After this initial condition subjects were asked to vocalise any covert rehearsals that they had been making in the first condition. Having the two conditions allowed the use of such a natural definition of rehearsal and to allow a comparison between the two conditions. Olson, Duffy and Mack (1981) warn that not all subjects are able to provide informative data and this is a possibility we intended to look for.

4.2 Experiment IX

4.2.1 Method

Design

The design of the texts was the same as in Experiment II. Each subject saw 40 texts which were split into four sessions of ten texts which will be referred to as *Blocks*. Mode was crossed with Block, and Matchtype was balanced by subject. The sequence of Mode and Matchtype was randomised.

The full design consisted of the following factors (and levels):

1. Mode (10 levels) of a text,
2. Individual (2 levels) within a text,
3. Property (4 levels) within a text,
4. Block (4 levels) of texts,

All factors were within-subject factors and were fully crossed. The reading time and any vocalizations (which were presumed to be overt rehearsals) for each sentence were recorded as was the recall for each text.

Materials

The vocabulary used in the experiment appears in Appendix C.1. The vocabulary set contained 48 words which was split into four groups of twelve each denoting occupation, nationality, stature and temperament and the four groups are referred to as dimensions. The generation of texts was the same as in Experiment II.

Apparatus

The materials were presented on a BBC model B microcomputer which recorded the times. A video recorder with microphone was used to record the computer's output to its monitor and the subject's vocalizations. This was to facilitate the logging of the vocal data by synchronizing it with the materials.

Procedure

Each subject was tested individually with the four blocks taking between 40 minutes to just over an hour to complete. Seventeen volunteer subjects took part in the experiment. They were provided with written instructions which were supplemented with detailed verbal explanation. Each text was presented to the subjects a sentence at a time. Subjects were instructed to read the texts as quickly as possible, consistent with recalling the individuals accurately.

The procedure within a block was the same as in Experiment I.

At the end of block one subjects were requested to vocalize any rehearsals they were making for the next two blocks. For the final block subjects were given the option of either continuing to rehearse aloud or of reverting to their original strategy (if this was different from overt rehearsal). It should be pointed out that subjects were given no instructions regarding particular rehearsal strategies.

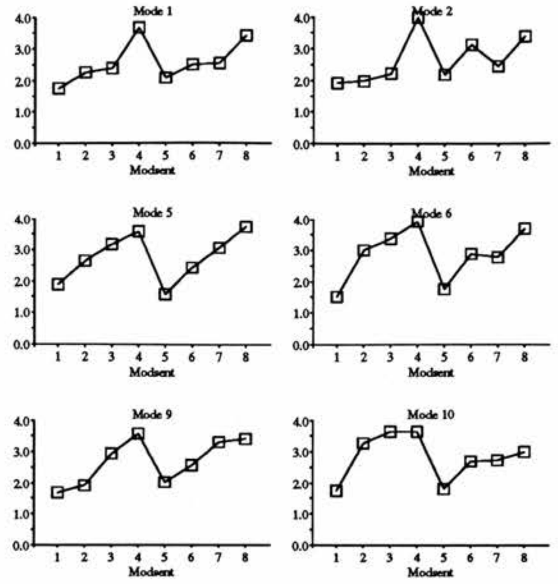

Figure 4.1: Mean reading times (sec) as a function of Modsent and Mode.

4.2.2 Results

Reading times and any overt rehearsal for each sentence were recorded as was the recall for each text. Two subjects out of the 20 subjects made no overt rehearsals in block four. This indicates that the majority of the subjects did not find the act of externalizing their rehearsal a processing burden. There is also the possibility that subjects found externalizing their rehearsal an advantage.

Reading Times

An analysis of variance was carried out, with subjects as the random factor, and mode, individual, property and block as fixed factors. There was a main effect of mode ($F(9,144) = 2.33, p < 0.02$), (see Figure 4.1 for the means of each mode by Modsent).

There was a main effect of property ($F(3,48) = 39.67, p < 0.0001$). This effect was due to a rise in reading time from properties one to four (the means were 1.79, 2.62, 2.88, 3.60 seconds respectively). There was a main effect of Block ($F(3,48) = 5.38, p < 0.004$) (the means were 2.59, 2.88, 2.91 and 2.51 for blocks one through four). There was no main effect of individual ($F(1,16) = 0.830, p = 0.38$).

The interaction between mode and property was significant ($F(27,432) = 2.98, p < 0.0001$), as was the interaction between individual and property ($F(3,48) = 7.07, p < 0.0001$) and also the interaction between block and property ($F(9,144) = 5.18, p < 0.0001$).

There were two significant three-way interactions one between mode, individual and property ($F(27,432) = 2.49, p < 0.0001$), and the other between block individual and property ($F(9,144) = 2.53, p < 0.02$).

The main effect of block indicates a task effect and the main effect of property shows the

Block	Cued	Bestfit
1	6.00	7.25
2	6.14	7.18
3	6.32	7.45
4	6.17	7.21

Table 4.1: The mean cued and bestfit recall scores by block (maximum score possible is 8 and cell frequency is 170)

semantic ordinal effect (SOE) described in Stenning, Shepherd & Levy (1988). The main effect of mode shows the importance of the pattern of switching reference unpredictably and has been found previous studies using the MIT (Experiments I and II).

The interaction between mode and property shows that the reading time for the referenced individual depends on, not only, what has already been learnt about it but also on the pattern in which the information has been presented. The interaction between block and property is inevitable if vocalizing simply takes longer than rehearsing covertly. The rehearsal processes may be the same in all four blocks but there appears to be a difference between block one and the other three because of the extra time taken to externalise any rehearsal. This explanation may also explain the three way interaction between mode, individual and property.

The importance of the interaction between individual and property taken with the interaction between mode, individual and property shows that although the increase in reading time with properties known of the referenced individual appears in this data, it is not completely independent of the temporal sequence of attributions to the two individuals. The obvious explanation is the uneven distribution of processes of referential change in the different modes.

Recall

Table 4.1 shows the mean recall accuracy (Cued and Bestfit) for each block.

The accuracy across blocks compares well with the accuracy found in Experiment I and shows that subjects were performing the task adequately in spite of the different sets of instructions. A one-way ANOVA on both types of scores shows an insignificant difference between blocks. For cued scores, $F(3,676) = 0.538, p > 0.05$ and for bestfit scores, $F(3,676) = 1.642, p > 0.05$.

Table 4.2 shows the two recall scores for each mode by conflict and agreement.

The pattern of confusions clearly replicates the pattern of findings in Experiment I for modes 1 to 7. It was predicted that mode 10 in Experiment II would exhibit a similar pattern of confusions as the other Modegroup 2 modes which it did in the free recall paradigm used. In this experiment which used the same cued recall procedure as Experiment I mode 10 clearly showed that there was a change of status. Mode 5 does show a reduced number of confusions compared to Mode 5 in Experiment I but does show a similar pattern to Mode 5 in Experiment II so it seems likely that the addition of the three new modes (8–10) causes some change in the rules for primary/secondary status assignment.

	Where criteria agree		Where criteria conflict		
Mode	Frequency	Cued=Bestfit	Frequency	Cued	Bestfit
1	58	7.16	10	3.30	7.30
2	61	7.20	7	2.43	6.86
3	40	7.20	28	3.25	7.50
4	59	7.25	9	2.44	6.78
5	56	7.28	12	2.83	7.17
6	36	7.11	32	3.09	7.34
7	30	7.07	38	2.95	7.63
8	58	7.33	10	2.70	7.50
9	66	7.32	2	5.00	8.00
10	41	7.29	27	3.26	7.30

Table 4.2: Mean recall scores by mode and by agreement *vs.* conflict of scoring criteria

Overt rehearsal protocols

Pre-processing and logging subjects' verbalizations The subjects verbalizations were not restricted in any way. However, the most relevant aspects of their protocols (within this experiment) was their use of the predicates, presented to them in the experimental texts. Therefore the aim of the logging process was to record any mention of an experimental predicate within a subject's verbalizations. A group of predicates associated with a sentence was called a *rehearsal commentary* and groups of predicates referring to the same individual within a rehearsal commentary was called a *segment.*

Logging and pre-processing For each sentence of a text there was an associated opportunity to rehearse which was taken to be the time that that sentence was on the screen. Any predicates mentioned within a sentence's opportunity to rehearse were associated with that sentence and made up that sentence's rehearsal commentary.

Each rehearsal commentary was divided into segments. These divisions were made according to the reference of the predicates within that rehearsal commentary. A predicate could refer to either individual or to both individuals described by the text. A segment started at the beginning of a rehearsal commentary and finished when there was a change of reference in which case a new segment would be started. A segment could not contain any repetitions of a particular predicate. As the predicates were logged their reference was recorded because it would not always be possible to infer a predicates reference at a later stage. Therefore the logging process contained an element of pre-processing in that subjects' verbalizations were split into commentaries and into segments each with their associated reference.

For example, if a person said "There is a bishop. There is a French bishop." then this would be logged as two segments, *bishop and french bishop*, with the same reference. If a person said "The bishop is French and the other one is French too." then this rehearsal commentary would be be logged as two segments, *bishop french and french* with different reference.

Rehearsal commentary strategies Table 4.3 shows the percentage number of segments per rehearsal commentary by subject for blocks one to four.

	number of segments				
Subject	0	1	2	3	4
1*	0	159	69	11	1
2*	8	186	39	6	1
3	123	113	4	0	0
5	0	220	19	1	0
6	148	88	4	0	0
7*	40	148	32	20	0
8	0	223	16	1	0
9	0	239	1	0	0
10	116	122	1	1	0
12	22	214	4	0	0
13*	76	77	81	6	0
14	114	124	2	0	0
16*	3	116	99	16	6
17	9	222	8	1	0
18*	5	114	115	5	1
19	120	106	9	4	1
20*	42	140	55	3	0
Totals	826	2611	558	75	10

Table 4.3: Frequency of commentaries by number of segments per commentary and by subject for sessions B, C, D. * denotes Subject Group 2. The total number of commentaries for each subject was 240.

Only the commentaries that contain predicates which correctly refer to one or other individual in the associated text have been included in Table 4.3. This is because incorrect and dual reference predicates are special cases and may need to be treated separately. Dual reference segments are segments which refer to both individuals and were prefixed by the subject with the word "both". They are special cases because it is unclear how to compare the number of properties rehearsed with the number of properties known. For example, if one of the properties in the segment is a property which is matched, then does the subject know one or two properties? If the properties ascribed to both individuals by the subject correctly describes only one of the individuals is the segment to be treated as incorrect? Incorrect segments are also special cases because there is no way of comparing the number of properties to the number of properties known. For these reasons segments which refer to both individuals and which are incorrect have been omitted from the analyses and the tables. The incorrect segments constitute 0.9% and the dual-reference segments constitute 2.2%. It is clear that for almost all subjects, if they said anything within an opportunity to rehearse then it is most likely that their rehearsal commentary contained one segment. However, there is a group of subjects, denoted by a * in the table, for whom the frequency of their two segment commentaries is greater than that of the remaining subjects. The group of subjects who are more frequently produce two-segment rehearsals will be referred to as Group 2 subjects and the rest of the subjects will be referred to as Group 1 subjects.

Table 4.4 shows the mean sentence-reading times for the two groups of subjects.

The mean time for Group 2 subjects is significantly longer than the mean time for Group 1 subjects ($t(3003.8) = -9.08, p < 0.01$). The recall scores for both groups are presented in Table 4.5 by cued and by best-fit criterion. The Table shows that the two groups of subjects

Group	Freq.	Mean	Std. dev
1	2400	251.87	176.63
2	1680	312.36	229.03

Table 4.4: Reading times (in centiseconds) by Group for sessions B,C and D.

		Cued		Bestfit	
Group	Freq	mean	std. dev	mean	std. dev
1	300	5.90	2.39	7.22	1.19
2	210	6.74	1.96	7.41	1.13

Table 4.5: Cued and Bestfit recall scores by Group for sessions B, C, D.

we equally good at correctly grouping the predicates together ($t(454.7) = -1.49, p > 0.1$) and that Group 2 subjects were better than Group 1 subjects at observing the cue or being able to identify which individual was mentioned first and which second ($t(492.5) = -4.21, p < 0.0001$).

The ten modes used in this experiment have been allocated to two groups referred to as Modegroups. Modes belonging to Modegroup 2 seem to result in a depressed recall performance which has been attributed to a confusion of the temporal identity of the two individuals (see Chapter 2). Modes belonging to Modegroup 1 do not exhibit this phenomenon. Table 4.6 shows that Modegroup has no effect on the number of segments per commentary for Group 1 subjects.

However, Group 2 subjects show an increased likelihood of producing two segment rehearsal commentaries in Modegroup 2 texts. This increase in likelihood is related to those sentences after a change of status in Modegroup 2 texts, as Table 4.7 shows.

An ANOVA was done for all subjects with subjects as a random factor and change of reference as a fixed factor (whether before or after a change of status) and number of segments as the dependent variable. It showed that the average number of segments per commentary was greater after a change of status (means were, 0.896 and 1.131 for before and after), $F(1,16) = 18.257, < 0.002$.

A rehearsal commentary which contains two segments about different individuals can be in two orders. The order of the segments can be described in several ways depending on how the indi-

Group	Modegroup	number of segments				
		1	2	3	4	Total
1	1	782	8	1	0	791
	2	802	11	3	0	816
	Total	1584	19	4	0	1607
2	1	582	154	17	3	702
	2	391	290	24	1	706
	Total	919	444	41	4	1408

Table 4.6: Frequency of commentaries by number of segments per commentary and by Modegroup and Group. Only commentaries from sessions B,C and D are included if they contain only segments which are correct, refer to one individual and do not refer to an individual already referred to by another segment in the same commentary.

Change of status	number of segments 1	2	3	4	Total
Before	170	32	9	0	211
After	221	258	15	1	495
Total	391	290	24	1	706

Table 4.7: Frequency of commentaries by number of segments per commentary before and after a change of status for Group2 Modegroup 2. Only commentaries from sessions B,C and D are included if they contain only segments which are correct, refer to one individual and do not refer to an individual already referred to by another segment in the same commentary.

Change of status	First individual referred to	Number of segments 1	2	3	4	Total
Before	Individual 1	103	28	0	0	131
	Individual 2	67	4	9	0	80
	Total	170	32	9	0	211
After	Individual 1	98	239	1	1	339
	Indidivual 2	123	19	14	0	156
	Total	221	258	15	1	495

Table 4.8: Frequency of commentaries by number of segments per commentary before and after a change of status and by order of individuals for Group2, Modegroup2. Only commentaries from sessions B,C and D are included if they contain only segments which are correct, refer to one individual and do not refer to an individual already referred to by another segment in the same commentary.

viduals are categorized; referenced and non-referenced, primary and secondary or introduced first and introduced second. Table 4.8 shows that after a change of status Group 2 subjects reading Modegroup 2 texts more frequently produce two-segment rehearsal commentaries in order of introduction than in counter-introduction order.

This difference also appears before a change of status but the total number of two-segment rehearsal commentaries is smaller and so the difference may be less important. The observation that the frequency of two segment rehearsal commentaries in introduction order is affected by a change of status suggests that the order retaining properties of rehearsal may be an aid to retaining the sort of information necessary to follow the recall cue.

Segment strategies In order to investigate the segment strategies each segment was categorized. Essentially this categorization depended on which predicates out of the known predicates appeared in a segment. There were five categories which applied to segments that contained correctly grouped properties *i.e.*, the properties of one of the individuals. (Category names are in parentheses)

1. The segment contained all the predicates known of the individual at that point in the text (Everything known).
2. The segment contained only the new property which was currently on the screen (New property only).

Sequence	N	Percentage
Misc.	98	2.6
Natural	3241	84.5
Presentation	397	10.4
Natural+sw	99	2.6
Total	3835	100.0

Table 4.9: Frequency of segments by sequence. Only segments from session b,c and d are included if they refer to the referenced individual only. The Natural+sw category referes to those segments in which the last two dimensions were in the Presentation order as opposed to the Natural order

3. The segment contained the two properties on the screen *i.e.*, the introducer and the new property. This category was included primarily to see if subjects were simply reading the current sentence off the screen (On screen properties).

4. The segment contained properties of the individual which had already been discovered in previous sentences (Everything learnt before).

5. If a segment did not fit into any of the other categories it was classed as miscellaneous (Misc.).

It should be noted that segments which refer to the non-referenced individual, that is, segments which are about the individual which is not referred to by the current sentence, can only fall into the miscellaneous and "everything known" categories. This is because, early on in texts a situation arises in which the categorization of a segment may be ambiguous. For example, if the current sentence is "There is a French bishop." and a segment contains *bishop French* then the segment contains properties which are on the screen and are all the properties known about the individual. In these cases the segment is assumed to belong to the "everything known" category because subjects are likely to say all that they know about an individual (3127 out of 3835 segments contain all the properties known). This means that the "properties currently on the screen" category may be a rather conservative estimate of the frequency of this category.

The sequence of the properties within segments has been categorized and is displayed in Table 4.9.

The "Natural" category refers to those segments in which the sequence of properties follows the so-called Natural order. This ordering arises when an individual is described in one sentence with the smallest number of words. For example, if one of the individuals described by a text is a bishop, who is French, meek and happy then this individual can be easily described as "A happy, meek, French bishop" which is shorter than saying "There is a bishop who is French, meek and happy". The sequence in which the properties of an individual has been learnt is called the Presentation sequence because the order of the properties is the same as their order of presentation. If a segment contained at least three properties and the properties belonging to the last two dimensions were in the presentation order but the segment was otherwise in the natural order then this it was classified as being in the natural order with the last two dimensions switched. If the sequence of properties within a segment fall into neither

Stratgey	Switch of reference to		No switch of reference from		Total
	Primary	Secondary	Primary	Secondary	
Incorrect	8	8	2	10	28
New property only	10	18	50	93	171
On screen properties	68	76	42	81	267
Everything learnt before	8	14	25	7	54
Misc.	27	11	32	47	117
Everything known	262	705	1070	397	2434
Total	383	832	1221	635	3071

Table 4.10: Frequency of segments by Strategy and by switch or continuation of reference. Only segments from sessions B,C and D are included if they refer to the referenced individual only.

of the above categories then it is classed as miscellaneous. The table clearly shows that most segments are in the Natural order.

Table 4.10 shows the frequency of segments by strategy, continuity of reference and status.

It is clear that when there is a switch of reference to the primary individual or a continuation of reference to the secondary individual that subjects do not rehearse all the properties they know of an individual. The table also shows that the properties they miss out are old properties or properties which they have learnt in previous sentences. This is to be expected because subjects must rehearse new properties if they are reading them (Baddeley, 1986) but they do not need to rehearse properties they already know.

An ANOVA was done with subjects as a random factor, object status and switch of reference as fixed factors with the number of properties of the referenced individual missed out divided by the number of properties known of the referenced individual as the dependent variable. Object status had a main effect (means 0.11 and 0.87 for primary and secondary respectively), $F(1,16) = 4.794, p < 0.05$. There was an interaction between object status and switch of reference, $F(1,16) = 16.556, p < 0.001$. Figures 4.2 shows the means.

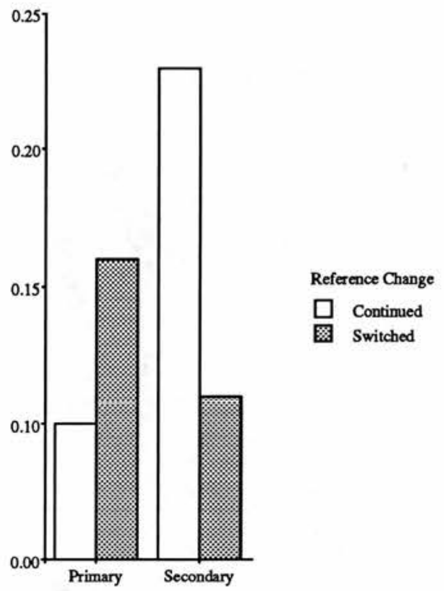

Figure 4.2: Number of properties missed out divided by number of properties known by status and switch of reference for blocks 2 to 4

4.3 Discussion

The results show that the request to subjects to externalise their rehearsal did not disrupt their performance on an MIT. Only two subjects avoided rehearsing overtly in the last session, the recall accuracies were the same across the four sessions and the reading time and recall results replicated previous findings (Experiment I and Experiment II). The reading time results showed the SOE and the effects of mode. The recall results replicated the results from Experiment I and confirmed the predictions about Modes 8 to 10. The pattern of recall also agreed with the pattern in Experiment II although the nature of the recall task was slightly different.

Although the general results show that subjects were performing in the usual way the analysis of the overt rehearsal commentaries does show a set of interesting individual differences. Group 2 subjects use more double segment rehearsals than Group 1 and this seems to enable them to recall the order of introduction much better than Group 2 subjects although the two groups are equally good at binding the properties together. Moreover this two segment strategy clearly appears to be related to sentences after a change of status which suggests that subjects are using two segment rehearsals to maintain order of introduction information and in fact the vast majority of the commentaries contain segments in introduction order. This is no great surprise if rehearsal is being used for its order maintaining properties and a change of status does disrupt order information.

The analysis of the segments related more closely to the regression model effects than did the rehearsal commentaries. Clearly the predominant strategy was to rehearse all the properties known about the referenced individual and this was reflected in the regression models as a positive coefficient for the number of syllables in the referenced individual's description.

The analysis of the rehearsal commentaries looked at the order in which the individuals were mentioned and the segments analysis investigated the order of the properties within a segment. The analysis showed that subjects preferred to rehearse the properties in the natural order although there were a few segments rehearsed in the presentation order.

The segment analysis also related directly to the negative word length effect when there is a switch of reference to the primary individual. The regression model showed that there was an acceleration in reading which was predicted by the number of syllables in the non-referenced individual's description. This acceleration could have been achieved by simply making the cognitive machinery run faster or some usual part of the processing could have been abbreviated. This analysis shows that when subjects switch reference to the primary individual they are likely to avoid rehearsing properties of the primary individual. Presumably this abbreviation accounts for at least part of the negative coefficient in the regression models. The analysis also shows that properties are missed out when subjects are processing a run of reference to the secondary individual. This agrees with Model 1 which has a coefficient with a negative value when there is a run of reference to the secondary individual.

These effects suggest a possible descriptive model of the relation between two stores, one which has phonological properties (P) and one which has semantic properties (S). Suppose that the secondary individual's description must be processed as a whole whereas the primary individual's description may be processed one property at a time. If this were the case then as new properties of the secondary individual were learnt then the whole description would have to be transferred from P to S. However, as new properties of the primary individual were learnt each new one could be transferred from P to S on its own.

When there is a switch of reference from the secondary individual to the primary individual the description of the secondary individual must be transferred from P to S otherwise without maintenance the description will decay, moreover the whole description must be transferred. The description of the primary individual must also be activated but not necessarily rehearsed because it has a robust representation held in some semantic form. Therefore as the primary individual's description is loaded into P and rehearsed, the secondary individual's description is being written to S. Because the secondary individual's description is fragile its processing takes precedence over the primary individual and so some of its (the primary individual's) description may be missed out when rehearsal starts and the amount that is missed out will depend on the length of the secondary individual's description.

A similar situation apparently arises when there is a run of reference to the secondary individual. As new properties are learnt the complete description must be written to S from P. Because the whole description is being transferred there may not be sufficient time to rehearse the complete description hence properties are missed out. Properties can be missed out because the representations in P do persist for some time without rehearsal.

When there is a run of reference to the primary individual there is no need to miss out any properties because the secondary individual's phonological representation has already been processed when the switch to the primary individual was made. Presumably while the run continues new properties are written to the semantic store for integration one at a time which means that when there is a switch of reference to the secondary individual there is no rush to write the primary individual's description.

Although this outline is very brief it does suggest that perhaps the two individuals are updated in different stores and this is one manifestation of their asymmetry. The primary individual's description is updated in a semantic store while the secondary individual's description is updated in a phonological one. This might also indicate why the primary individuals shows more semantic effects like the number of properties in its description. Although the above model is speculative it does hilight the opportunity to study memory updating by investigating switching reference which will be returned to in Chapter 6. It should be noted that although the model is presented using serial-computer-related terms it is meant to be descriptive and does not contain any commitment to a particular architecture of cognition.

The subjects who took part in this experiment showed a range of individual differences on the task. One group of subjects predominantly rehearsed the descriptions of both individuals while the other group rehearsed only the description of the referenced individual. This is a difference which was not found in the Experiments I and II and probably reflects differences in the way people interpreted the instructions to rehearse rather then genuine differences in task performance. The differences between subjects may also be due to differences in their ability to supply the sort of information that was requested (Olson, Duffy & Mack, 1983). The ANOVA showed that the effect of missing out properties was evident in all subjects so the differences were likely to be due to interpretation rather than TOL ability.

This experiment clearly fulfilled its two aims. It revealed new aspects about the use of rehearsal and clarified the effects which were revealed by the regression modeling of the reading time data from Experiments I and II. Furthermore, the recall analysis showed that the pattern of confusions in Experiment I were robust and provided further support for the primary secondary distinction by confirming the predictions about Mode 10.

While the replication of the reading time and recall results showed that overt rehearsal did not disrupt subjects' performance of an MIT and was therefore a valid technique, it is still hard to determine exactly how rehearsal is being used. The analysis of the overt rehearsal implied that it was being employed by some subjects, at least, to help maintain order of introduction information or information about sequence. If this was the case then it would be expected that removing subjects' ability to rehearse during an MIT would have profound consequences. The next chapter describes an experiment which used articulatory suppression to investigate the use of rehearsal in an MIT.

Chapter 5

Articulatory suppression in complex and simple tasks

Summary

Two sources of evidence had suggested that phonological short-term memory was being used when reference changed unpredictability in texts which described individuals. As a way of gathering more evidence subjects were asked to read similar texts while repeating an irrelevant word which should block the use of phonological memory. The manipulation did not disrupt subjects use of phonological memory. This brought into question the theory which was used to explain phonological effects in short-term memory.

5.1 Introduction

THE EXPERIMENTAL evidence described in Chapters 2 and 4 appears to be consistent in supporting the view that articulatory rehearsal is used in the MIT. The results from Experiments I and II showed that the number of syllables in individuals' descriptions accounted for some part of the reading times. The regression coefficients suggested that subjects missed out properties when reference switched. Experiment IX directly recorded subjects' rehearsal and confirmed that subjects were missing out items from their rehearsal when they switched reference. Given the broad evidence base it seems reasonable to conclude that articulatory rehearsal is used in the MIT with unpredictable modes.

In spite of the convergent evidence described above there are still other effects which should be evident if articulatory rehearsal is really being used. For example, Baddeley, Thomson and Buchanan (1975) found that the word length effect was reduced, if not abolished, by articulatory suppression. There are other diagnostic effects like phonemic confusions (Baddeley, 1976; Baddeley & Hitch, 1974; Conrad, 1964, 1970; Wickelgren, 1969) and susceptibility to unattended speech (Salamé & Baddeley, 1982). However, because word length effects were interpreted as due to articulatory rehearsal then the effects should be reduced or abolished by articulatory suppression. Therefore, in order to increase the support for the articulatory rehearsal interpretation of the word length effects, an MIT was carried out under suppression conditions. First, such an experiment would test the interpretation of the word length effects and secondly by disrupting rehearsal, evidence should be found for the role of rehearsal in the MIT.

Several studies have attempted to investigate the effects of suppression on reading. Levy (1977) used counting aloud as a suppression task which did cause a decrement in subjects' ability to recognize sentences they had read. Levy (1978) used a task where subjects were required to make paraphrase judgements (recognize paraphrases of previously presented sentences) in which suppression had no effect although suppression did have an effect on verbatim recognition. Slowiaczek and Clifton (1980) also used counting and a second suppression task requiring the repetition of "Cola". They found that suppression did have an effect on written prose comprehension demanding the integration of concepts but not on the recall accuracy of single concepts.

Baddeley, Eldridge and Lewis (1981) observe that in the Levy, and Slowiaczek and Clifton work suppression does have an effect which is ambiguous. The effect of suppression could arise from the way in which comprehension was tested or the way in which comprehension was achieved. The problem essentially lies in the requirement for subjects to remember something about the stimulus and it is known (Baddeley, Thomson & Buchanan, 1975) that suppression affects verbal memory. Therefore, even if comprehension had been unaffected by suppression the memories of the stimulus material, although comprehended, could be impaired and so a misleading result produced.

Baddeley and Lewis (1981) attempted to use an on-line phonological effect which did not rely on verbal memory by developing work carried out by Baddeley and Hitch (1974). Baddeley and Hitch (1974) reasoned that if comprehension depended on a phonetic code then similarity among words in sentences should impair performance as it did in immediate memory span tasks. They presented subjects with sentences like (5.1a) and contrasted them with sentences like (5.1b) which were semantically equivalent but phonemically dissimilar. Subjects were required to judge whether each sentence was correct or not. Incorrect sentences were generated by swapping items in the sentences, like (5.1c). They found that subjects were slower at the task when presented with phonemically similar sentences compared to phonemically dissimilar sentences although phonemic similarity had no effect on error rate.

(5.1) a. Rude Jude chewed his crude stewed food.

b. Rough curt Jude ate his plain boiled meat.

c. Crude rude chewed Jude stewed food.

Baddeley and Lewis (1981) replicated the effect and investigated whether articulatory suppression (repetition of the digits 1 2 3 4 5 6) removed the phonemic similarity effect. They found that suppression did increase the number of errors but did not affect the size of the phonemic similarity effect. In summary, phonemic similarity had an effect on decision time but no effect on error rate but suppression had an effect on error rate but no effect on decision times. Therefore the role of phonological coding was not clear although it did seem that articulatory coding was involved.

Baddeley, Eldridge and Lewis (1981) noted that the materials used in Baddeley and Hitch (1974) and in Baddeley and Lewis (1981) did not represent a realistic sample of the materials used for normal reading. Therefore, they carried out an investigation using an anomaly detection task with long and relatively complex sentences. Subjects were presented with sentences like (5.2). An anomalous version was created by substituting a word like *rent* for *pain*. They also used a suppression task in which subjects were required to repeat the digits 1 2 3 4 5 6 at a rate of 4 digits per second.

(5.2) She doesn't mind going to the dentist to have fillings, but doesn't like the pain when he gives her the injection at the beginning.

Suppression had no effect on processing speed although there was a difference in speed between the consistent and anomalous conditions where anomalous sentences were processed faster. Suppression did affect accuracy, making subjects who were suppressing more likely to accept an anomalous sentence. They further controlled for the attentional affects of suppression by repeating the experiment with a tapping task. In this experiment subjects were required to say the word *the* at a rate of 200 per minute (3.33 per second) for the suppression task and to tap with their non writing hand at the same rate. They found consistent results (where suppression affected accuracy but tapping did not)which further demonstrated that the crucial factor in the suppression task was not due to attentional demands.

They conclude that articulatory coding is not crucial to comprehension (because suppression has no effects on speed of processing) but do suppose that it is useful in detecting subtle changes in the wording of a text where changes of order are involved.

Besner (1987) reviewed the use of articulatory suppression as a tool for investigating the role of phonological coding in reading and lexical access. He discusses several experiments which have examined the effect of suppression on rhyme judgement and concludes that their logic is flawed and that the results really show that there are possibly two phonological codes involved in reading: one code which may be affected by suppression, the other which need not necessarily be. Crucially, he observes that the code which underlies the phonemic similarity effect and the word length affect is affected by suppression. In addition, this result (Baddeley, Thomson & Buchanan, 1975) has been replicated several times (*e.g.*, Besner & Davelaar, 1982; Coltheart, Avons & Trollope, 1990).

Although suppression may not affect all phonological coding it does seem clear that it does affect the coding which underlies word length effects. Therefore, suppression should be a reliable test of the word length effects found in the MIT. Consequently, an experiment was designed using a simple MIT with several extra conditions. Following Baddeley, Eldridge and Lewis (1981) there were two secondary tasks: one was an articulatory suppression task and the second was an attentional control in which subjects were required to tap at the same rate. It was predicted that the word length effects would be abolished by the suppression task and that the tapping and normal conditions would still show word length effects. Normal is used to refer to the condition in which subjects were not required to perform a secondary task which was the "normal" or usual condition the MIT is performed under.

An extra factor was added which was designed to investigate the role of the articulatory loop in the MIT. As discussed earlier, it appears that subjects employ a representation strategy which depends on the encoding or use of serial order/sequence. This strategy is particularly important when the modes used in the MIT are unpredictable. The discussion in the articulatory loop literature makes several mentions of the supposed property of the loop that it is used to retain order information. Furthermore, Baddeley, Vallar and Wilson (1987) make this claim explicit in a a study of two people with brain damage. Therefore, if suppression really disrupts the loop then sequence information will be lost and if sequence information is used in unpredictable modes then suppression will have an effect on them, in contrast, to a negligible effect on predictable modes. Order information may be used in predictable modes but the point is that the mapping from order to binding is not disrupted.

Mode	Order of Presentation							
0	A1	B1	A2	B2	A3	B3	A4	B4
1	A1	A2	A3	A4	B1	B2	B3	B4
2	A1	A2	A3	B1	A4	B2	B3	B4
3	A1	A2	B1	B2	B3	B4	A3	A4
4	A1	B1	A2	A3	B2	A4	B3	B4
6	A1	B1	B2	B3	A2	A3	B4	A4
7	A1	B1	B2	B3	B4	A2	A3	A4
8	A1	A2	B1	A3	A4	B2	B3	B4
9	A1	A2	B1	B2	A3	A4	B3	B4
10	A1	B1	B2	B3	A2	B4	A3	A4

Table 5.1: The ten modes of presentation used in Experiment X. A denotes the first individual, B the second individual, 1 the introducing dimension, 2, 3 and 4 the second, third and fourth property dimensions.

5.2 Experiment X

5.2.1 Method

Subjects

Forty eight volunteer subjects took part in three sessions and were paid £10. For each subject, no more than two sessions were performed in one day and were never concurrently.

Design and Materials

The standard MIT design was used for the construction of texts. Subjects read texts composed of eight sentences in a self-paced reading time task and recalled the two individuals which had just been described in a cued recall phase.

The novel part of the design was the manipulation of three variables: Task, Word Length and Predictability. Task referred to the secondary task which subjects would undertake while performing the primary task, namely the MIT. This variable had three levels (Normal, Suppression and Tapping) and each level of the variable applied to one session of 32 texts. Word Length referred to the number of syllables in the adjectives used to make a pair of individual's descriptions and had two levels: Short (1 syllable) and Long (2 or 3 syllables). Predictability referred to the mode of the text which had two levels: Predictable and Unpredictable. Predictable texts were either I×I (mode 1) or P×P (mode 0) texts and Unpredictable ones were in modes 2 to 10. See Table 5.1 for a full list of the modes.

Because Predictable and Unpredictable modes depended on their contexts (which other other modes they were presented with) they were grouped. So each session of 32 texts was split into two half-sessions of 16 predictable texts and 16 unpredictable texts. Each half session had 8 Short texts and 8 Long texts.

To counteract any effects of Task ordering or Predictability ordering two counterbalancing schemes were used. Table 5.2 shows the scheme for Task for groups of 6 subjects and Table 5.3

	Session Order		
Subject	1st	2nd	3rd
1	N	S	T
2	N	T	S
3	S	N	T
4	S	T	N
5	T	N	S
6	T	S	N

Table 5.2: Counterbalancing of Task Session Order across groups of six subjects (N = Normal, S = Suppression, T = Tapping).

	1st session		2nd session		3rd session	
Sub-session	1st	2nd	1st	2nd	1st	2nd
Subject 1	NP	P	P	NP	NP	P
Subject 2	P	NP	NP	P	P	NP

Table 5.3: Counterbalancing for the order of Predictable (P) and unpredictable (NP) blocks.

shows the scheme for Predictability counterbalanced across pairs of subjects. The two schemes were crossed so that Task and Predictability were counterbalanced across groups of 12 subjects.

It was important to make sure that Mode and Matchtype were crossed and counterbalanced across subjects and with Task and Word Length as well. When all four variables are crossed, there are 384 conditions (8 unpredictable modes, 8 matchtypes, 3 tasks and 2 word lengths) and each subject saw 48 texts so 8 subjects could see all the conditions ($8 \times 48 = 384$). The counterbalancing was achieved by choosing an arbitrary pairing between the 8 unpredictable modes and the 8 matchtypes for each combination of Task by Word Length within a subject. The matchtypes were rotated for each new subject until all 64 combinations had appeared which meant that a new arbitrary correspondence was generated for each group of 8 subjects. Table 5.4 shows an example of mode matchtype pairing for a particular combination of Task and Word Length.

The counterbalancing scheme has been described for unpredictable texts only and a similar scheme was employed for predictable texts but because there were only two predictable modes the number of combinations were reduced by a factor of 4 so all the combinations were exhausted for pairs of subjects.

	Subject							
Mode	1	2	3	4	5	6	7	8
1	7	1	8	5	4	2	3	6
2	6	7	1	8	5	4	2	3
3	3	6	7	1	8	5	4	2
4	2	3	6	7	1	8	5	4
5	4	2	3	6	7	1	8	5
6	5	4	2	3	6	7	1	8
7	8	5	4	2	3	6	7	1
8	1	8	5	4	2	3	6	7

Table 5.4: The scheme for rotating Mode/Matchtype combinations over subject for a particular Task/Word Length combination. Cell entries are Matchtypes.

This counterbalancing/crossing scheme meant that mode and matchtype were counterbalanced within a subject but not crossed and that Mode was crossed with Task and Word Length (within a subject) and that Matchtype was crossed with Task and Word Length (within a subject).

The recall phase of the task was cued (see Experiments I and IX). The cue was crossed with Mode and Task but was allowed to vary with Word Length at random.

The texts were generated in the same way as in previous MIT designs. The vocabulary was designed to control word frequency but because of other constraints the control was not complete. Appendix D.1 shows the vocabulary by dimensions and word length.

Procedure

The procedure for each text was the same as the standard MIT procedure described in earlier experiments with cued recall. However, the procedure was supplemented by a secondary task in two of the sessions: Suppression and Tapping. While a text was being read subjects were asked to repeat the word "the" at a rate of 2 words per second in the Suppression condition or use their free hand to tap the desk at a rate of 2 taps per second in the Tapping condition. They started the secondary task when the setting appeared and stopped when the question appeared. Before each session the subject practised the secondary task with a metronome and under the supervision of the experimenter. The subject then did three practice trials, again under the supervision of the experimenter. Once the practice was over and the instructions had been understood the experimenter left. However, subjects were warned that the experimenter would listen to them at random intervals and that they would be prompted if they were not performing the secondary task at the correct rate of two actions per second. The subjects were told that the session would be made up of 16 predictable texts and 16 unpredictable texts and they were informed prior to each half-session what the form (whether predictable or unpredictable) of the following texts would be. Two of the practice trials were predictable texts and the remaining one was unpredictable and subjects were forewarned of these differences.

5.2.2 Results

Analysis of the results showed that there were no interactions between Task and Word-length which would be expected if Suppression was really disrupting articulatory rehearsal. Therefore the complete analysis contains many subsidiary analyses which were conducted to ensure that no interaction was missed and that the experiment was correctly conducted and related to previous experiments with the MIT. Rather then report all the analyses here they are fully described in Section D.2 in Appendix D.

Given the lack of an interaction between Task and Word-length there were two possible explanation of this finding. First, it could have been simply that subjects were not rehearsing the suppression item enough and were allowing themselves to rehearse the primary material. Second, the materials could have had some lexical property which correlated highly with word length and was unaffected by suppression. The next experiment describes an investigation of

Baddeley		MIT							
		Cohort 1		Cohort 2		Cohort 3		Cohort 4	
short	long	short	long	short	long	short	long	short	long
stoat	puma	king	actress	dutch	scottish	fat	hungry	rich	graceful
mumps	measles	queen	bishop	thai	english	thin	thirsty	poor	awkward
school	college	boy	soldier	swiss	russian	tall	quiet	good	peaceful
greece	peru	girl	sailor	czech	american	short	noisy	bad	violent
crewe	exeter	maid	mother	greek	spanish	strong	generous	sane	gentle
switch	radio	chef	daughter	basque	italian	weak	modest	mad	heavy
maths	botany	clerk	brother	french	danish	young	alert	shy	healthy
maine	utah	judge	sister	welsh	swedish	old	innocent	gruff	delicate

Table 5.5: Pools of words used in Experiment XI by vocabulary set (Baddeley and MIT which is also subdivided by Cohort) and by Word length.

the second alternative.

5.3 Experiment XI

The aim of this experiment was to compare the materials used in the last experiment (Experiment X) with materials used by Baddeley and others (Baddeley, Thomson & Buchanan, 1975; Coltheart, Avons & Trollope, 1982) to demonstrate the effect of suppression on word length effects.

5.3.1 Method

Subjects

Twenty postgraduate students and members of staff of University of Edinburgh acted as subjects and were paid £3 for their participation.

Materials

Two vocabulary sets were used to make up the materials. One set was the vocabulary used in Experiment X (MIT vocabulary) and the other set was made up from the pool of words used in Baddeley *et al.* (1975). The MIT vocabulary is made up of four cohorts and each cohort contains eight short words (monosyllabic) and eight long words (di- and trisyllabic). The words taken from Baddeley et al.'s experiment were chosen so as to mirror the ratio of word lengths in the MIT vocabulary: Eight words were taken from their pool of ten monosyllabic words to make the pool of short words and five words were taken from their disyllabic pool and three words from their trisyllabic pool to make up the pool of long words. The reason three words were taken from their trisyllabic pool was to reflect the ratio of di- to tri- syllabic words in the MIT vocabulary. See Table 5.5 for a complete list of the vocabulary items.

Sixteen lists of six short words and sixteen lists of six long words were drawn from the two pools made from Baddeley et al.'s materials. For the MIT vocabulary, four lists of six short words and four lists of six long words were drawn from the pools of words for each cohort. Six items (one more than the list length in Baddeley, Thomson & Buchanan, 1975) were used to make each list to ensure that the total word length of any list was large enough to ensure some inaccuracies and avoid a possible "ceiling effect". The items were drawn from the pools randomly constrained in such a way that each item occurred equally often. None of the lists contained any repeated items and the order of the items was random. The order of the lists was also randomised. Each subject saw the same list.

A BBC model B microcomputer was used to present each item of the sixty four lists. The start of a list presentation was initiated by the subject, and each word was displayed successively for 1.5 sec. Words were displayed in the centre of the screen printed in upper-case letters.

Procedure

Subjects in the silent condition were told that sixty four lists of six words each would be presented on the screen, and that they would read each word as it was shown. At the end of each list a row of asterisks was shown, and this was the cue to begin written recall. They were required to write down the words in a row reflecting the serial order of the stimulus and indicating with a dash the location of any item not recalled. Subjects were given as long as the wished to recall the items before proceeding to the next trial. Before the main experiment started each subject had a practice of two lists during which the experimenter was present to answer any questions.

In the articulatory suppression condition subjects were asked to repeat "the" continuously at a rate of about two utterances per second. Before the experiment they practised this with a metronome. Subjects recalled the lists in silence and were instructed to start articulating before initiating the next list presentation.

5.3.2 Results

The number of items recalled in the correct serial position was calculated for each list for each subject. Analysis of variance with one between-subjects factor (suppression) and two within-subjects factors (vocabulary and word length) showed three main effects of suppression ($F(1,18) = 14.20, p < 0.002$), vocabulary ($F(1,18) = 14.65, p < 0.002$) and word length ($F(1,18) = 24.28, p < 0.001$). Words were recalled better when read in silence than in the articulation condition, MIT vocabulary was better recalled than Baddeley vocabulary and short words were better recalled than long words. There was an interaction between vocabulary and word length ($F(1,18) = 11.06, p < 0.004$) which showed that the word length effect was larger for MIT words than Baddeley words. The expected interaction between word length and suppression was also found ($F(1,18) = 5.02, p < 0.04$) which showed that the word length effect was greatly reduced under suppression.

A second analysis of variance was done on the data for the MIT vocabulary only with one between subjects factor (suppression) and two within-subjects factors (cohort and word length). This showed main effects of suppression ($F(1,18) = 14.53, p < 0.002$), word length

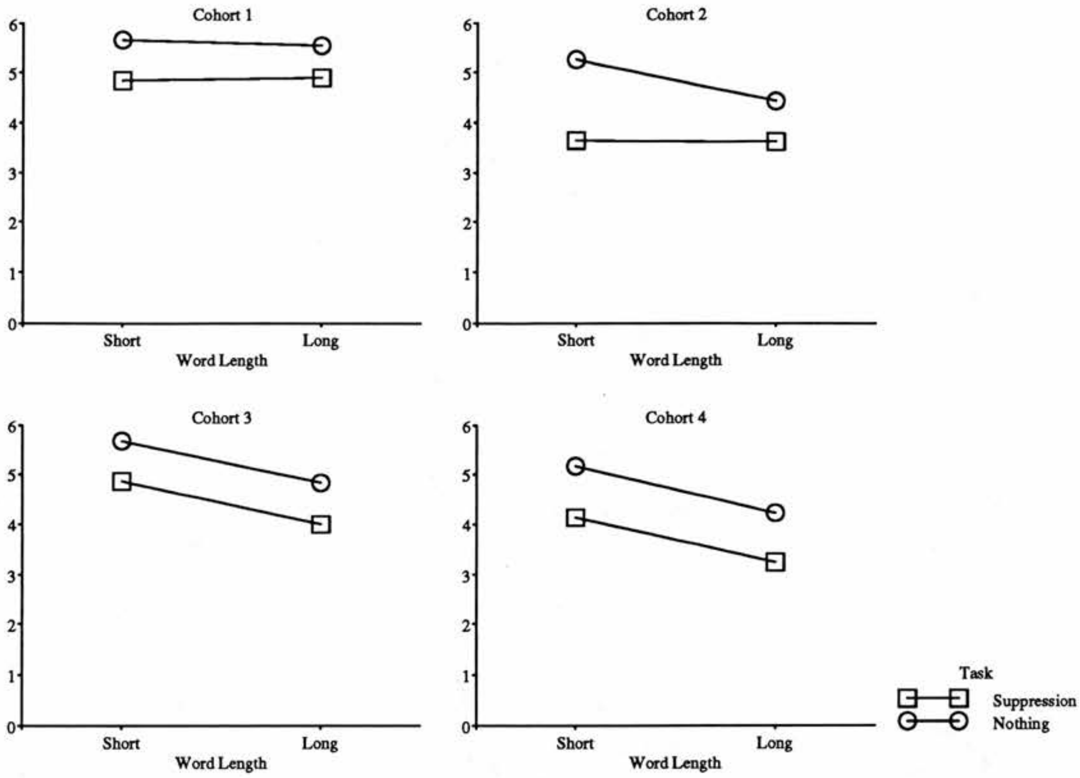

Figure 5.1: Mean recall accuracy out of 6 by Cohort, Secondary task and Word length for MIT vocabulkary only (N = 40).

($F(1,18) = 30.28, p < 0.0001$) and cohort ($F(3,54) = 30.28, p < 0.0001$). The effects of suppression and word length were the same as the previous analysis and the effect of cohort is clearly seen in the interaction between cohort and word length ($F(3,54) = 4.26, p < 0.01$). Figure 5.1 shows the interaction between suppression, word length and cohort. This interaction was not significant.

Vocabulary and frequency

The MIT vocabulary used in Experiments X and XI was not selected with an analysis of frequency in mind because it had already demonstrated in Experiment II that frequency could not account for the word length effects which we observed. However the well known correlation between word length and frequency may remain a candidate explanation of the lack of an interaction between suppression and word length.

An one-way analysis of variance for each cohort revealed the extent of the relationship between word frequency (Francis & Kucera, 1982) and number of syllables. For cohort 1 there was no effect ($F(1,14) = 0.777, p < 0.4$), no effect for cohort 2 ($F(1,14) = 1.39, p < 0.3$), no effect for cohort 4 ($F(1,14) = 1.084, p < 0.4$) and the effect of frequency was significant for cohort 3 ($F(1,14) = 9.17, p < 0.01$). It is clear from Figure 5.1 that cohorts 3 and 4 behave similarly with respect to word length and suppression and yet they are very different with respect to the relationship between frequency and word length. Cohorts 1 and 2 show an effect of suppression on word length and yet the relationship between frequency and word length is insignificant.

These two observations make it very unlikely that frequency can account for the inter cohort differences.

5.3.3 Discussion

The results show that the subjects who took part in this experiment were able to perform the articulatory suppression task adequately and that the materials used in Experiment X behaved differently from the materials used in Baddeley *et al.* (1975). It seems that word frequency is an unsatisfactory explanation of this difference and that the most likely explanation is concerned with the degree of relatedness between the words of a list.

5.4 Experiment XII

The last experiment (Experiment XI) did not monitor rate of articulatory suppression so there is still a possibility that rate of suppression varied with the two material sets. Furthermore even though frequency was considered an unlikely explanation it was tightly controlled in this experiment by introducing new vocabulary which maintained the same semantic properties of the materials used in Experiment X. A Tapping condition replaced the Normal condition because Experiment X had shown that Tapping produced similar effects to Suppression and it was necessary to show that the two conditions did have a different effect on word length. Also a Tapping condition is a better control than Normal because effects under Normal conditions have been widely reported and Tapping does at least impose an attentional demand which is closer to the attentional demands of articulatory suppression.

5.4.1 Method

Subjects

Twenty two postgraduate students and members of staff of University of Edinburgh acted as subjects and were paid £3 for their participation. Two of the subjects data was discarded because they showed a word length effect with improved recall accuracy for long words which was diagnostic of a non-articulatory strategy. Baddeley, Thomson and Buchanan (1975) report an imagery strategy which gave rise to this sort of word length effect.

Materials

In order to control for effects of word frequency in the MIT vocabulary an ANOVA was carried out on the word frequencies (number of texts out of 500 in the Kucera & Francis (1967) corpus) for the short and long groups for both vocabulary sets. For the Baddeley vocabulary set $F(1,11) = 0.796, p < 0.4$ and for the MIT vocabulary set, $F(1,18) = 0.766, p < 0.4$, Table 5.7 shows the number, means and standard deviations used in the analyses.

Vocabulary set							
Baddeley Words				MIT Words			
Short		Long		Short		Long	
STOAT	–	PUMA	–	FAT	37	GENTLE	25
MUMPS	–	MEASLES	2	THIN	62	HEAVY	86
SCHOOL	139	COLLEGE	78	STRONG	133	HUNGRY	17
GREECE	11	PERU	2	WEAK	23	THIRSTY	5
CREWE	–	BLACKPOOL	–	RICH	60	QUIET	59
SWITCH	18	KETTLE	3	POOR	75	NOISY	6
MATHS	–	PHYSICS	12	SANE	5	FRIENDLY	44
MAINE	9	UTAH	4	MAD	28	HOSTILE	17
SCROLL	–	ESSAY	19	SHY	13	MODEST	26
ZINC	5	CARBON	7	GRUFF	3	GENEROUS	23

Table 5.6: Vocabulary by pools and word length. The number after each word refers to the number of texts it occurs in out of 500 in the Kucera and Francis (1967) word count. Words which were not in the Kucera and Francis corpus are marked with a –

	Vocabulary set					
	Baddeley			MIT		
Word length	Number	Mean	SD	Number	Mean	SD
short	5	36	57	10	44	40
long	8	16	26	10	31	25

Table 5.7: Frequency (number of texts out of 500 in the Francis and Kucera (1967) corpus) means and standard deviations (SD) for the two vocabulary sets by word length

For each subject a set of forty lists was generated. Words from each of the four pools of words (see Table 5.6) made 10 lists. Each list was made up by sampling at random without replacement from one of the pools of words. The order of lists was random and the order of the words within a list was randomized. The only restriction on the generation was that each word occur an equal number of times for each subject. This was to ensure that a materials and subjects analysis would be possible.

A BBC model B microcomputer was used to present each item of the forty lists. The start of a list presentation was initiated by the subject, and each word was displayed successively for 1.5 sec. Words were displayed in the centre of the screen printed in upper-case letters.

Procedure

Subjects in the both conditions were told that forty lists of six words each would be presented on the screen, and that they would read each word as it was shown. At the end of each list a row of asterisks was shown, and this was the cue to begin written recall. They were required to write down the words in a row reflecting the serial order of the stimulus and indicating with a dash the location of any item not recalled. Subjects were given as long as they wished to recall the items before proceeding to the next trial. Before the main experiment started each subject practised reading and recalling three lists with the appropriate secondary task.

In the suppression condition subjects practised saying the sequence of numbers "1 2 3 4 5 6" at a rate of three items per second with the help of a metronome. In the tapping condition

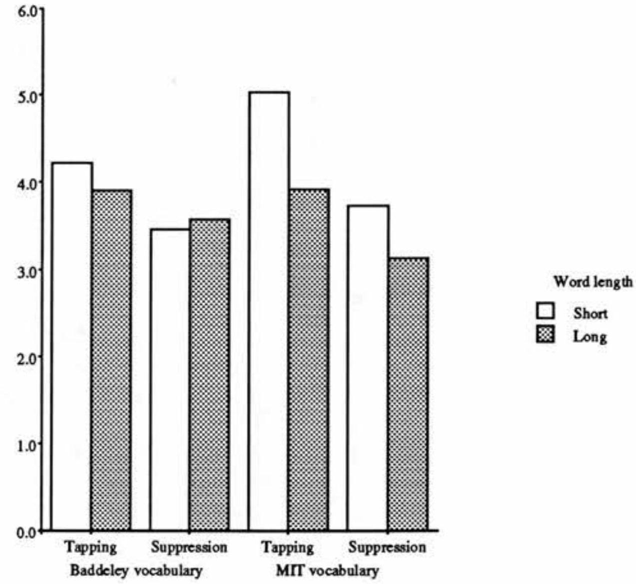

Figure 5.2: Mean recall accuracies out of six by task, word length and vocabulary

subjects practised tapping the desk at a rate of three taps per second with the help of a metronome. When the subjects (in either condition) had adjusted correctly to the speed then they proceeded to the practice trials and then to the main experiment. Subjects were instructed to start tapping or articulating before they pressed the space bar to initiate the presentation of a list and to stop when the line of asterisks appeared.

While the subjects tapped or suppressed the experimenter remained in the room to record the frequency of their secondary responses. If a subject fell below 2 and a half responses a second or rose above 3 and two thirds of a response per second then the subject was prompted before the next trial began.

5.4.2 Results

Analysis of variance was carried out with subjects as a random factor, task as a fixed between factor and word length and vocabulary as fixed within factors. There was a main effect of task ($F(1,18) = 4.5, p < 0.05$) and of word length ($F(1,18) = 45.24, p < 0.0001$). There were three interactions: between task and vocabulary ($F(1,18) = 5.38, p < 0.04$), between word length and task ($F(1,18) = 11.19, p < 0.004$) and between vocabulary and word length ($F(1,18) = 40.63, p < 0.0001$). Figure 5.2 shows the mean accuracies by task, word length and vocabulary.

Simple Effects

The interaction between task and word length was tested at both levels of the vocabulary variable. The interaction was significant at both: $F(1,18) = 5.33, p < 0.04$ for the Baddeley vocabulary and $F(1,18) = 7.86, p < 0.02$ for the MIT vocabulary. In order to determine the nature of these results four more simple effects tests were done to investigate the effect of

Vocabulary	Task	$F(1,18)$	p
Baddeley	Tapping	5.54	< 0.04
Baddeley	Suppression	0.83	> 0.3
MIT	Tapping	72.93	< 0.0001
MIT	Suppression	20.93	< 0.0003

Table 5.8: Simple effects of word length by task and vocabulary

Vocabulary	Word length					
	Short			Long		
	N	Mean	SD	N	Mean	SD
Baddeley	5	19.42	2.03	5	19.85	1.35
MIT	5	16.68	1.76	5	18.52	1.24

Table 5.9: Mean, Number (N) and standard deviation (SD) of time to say in seconds for the two vocabularies and two word lengths

word length at the four combinations of the vocab and task variables. Table 5.8 shows that although both interactions were significant they are different in character. For the Baddeley materials the word length effect has been abolished by suppression but a word length effect remains for the MIT materials under suppression although the size of the effect is less than the size of the effect under Tapping.

Time to say effects

It is clear from Table 5.2 that under tapping conditions the two vocabularies behave differently or rather the short words in the two vocabularies are different. This might explain the different behaviour of the two vocabularies. Time-to-say data showed that there was a large difference in time-to-say between the two sets of short words. Following Baddeley, Thomson and Buchanan (1975) the four lists were made up from each of the four pools by taking the ten words in each pool, randomizing the order, and repeating each one five times. This was to ensure that the maximum number of identical consecutive words was two. Then five subjects were asked to read each of the four lists twice and the time they took in seconds was recorded using a hand-held stopwatch. Table 5.9 shows the mean time to say in seconds for the lists. A two way ANOVA on the data showed only a main effect of vocabulary ($F(1,4) = 13.52, p < 0.03$) although the interaction between vocabulary and word length was approaching significance, $F(1,4) = 3.14, p < 0.2$. Although the apparent effects in the data were not significant the number of subjects was small and it is well known that there is a lot of between subject variability in time-to-say measures. Unfortunately we were not able to recall the original subjects so this data was only a *post hoc* attempt to investigate the apparent difference between the two vocabularies.

It may appear that it is the difference in the time-to-say of the short words in the two vocabularies which is causing the difference between the vocabularies interactions with suppression. However it is clear that although the long words in the two vocabularies have similar recall accuracies under tapping ($F(1,18) = 0.01, p > 0.9$) they do diverge significantly ($F(1,18) = 11.48, p < 0.004$) under suppression. This implies that the the differences between the two vocabularies does not entirely depend on differences in time-to-say.

Word length	Number	Mean	SD
Short	400	26.92	2.47
Long	400	26.69	3.01

Task	Number	Mean	SD
Suppression	400	25.83	2.84
Tapping	400	27.79	2.27

Table 5.10: Means for task and word length for secondary tasks

Materials analysis

An ANOVA was carried out on the vocabulary item as a random factor instead of subjects which meant that word length and vocabulary both became fixed between variables and task remained a fixed within factor. The dependent variable was the probability of correct recall for a particular item. The results clearly reflected the pattern of results in the subjects analysis. There were significant main effects of task ($F(1,36) = 101.05, p < 0.0001$) and word length ($F(1,36) = 32.34, p < 0.0001$). Again all the two-way interactions were significant: task by vocabulary ($F(1,36) = 10.13, p < 0.004$), task by word length ($F(1,36) = 8.96, p < 0.006$) and word length by vocabulary ($F(1,36) = 20.75, p < 0.0002$). Of course, the means by task, word length and vocabulary are all a sixth of the ones reported in Table 5.2.

Analysis of secondary task measures

An ANOVA was carried out on the secondary task data with subjects as a random variable, task as a between subjects fixed variable and vocabulary and word length as fixed within variables. The dependent variable was the number of secondary task responses (taps or items uttered) made during the nine seconds each trial lasted. There was one effect which was significant and one effect which approached significance: task ($F(1,18) = 4.60, p < 0.05$) and word length ($F(1,18) = 3.74, p < 0.07$). Table 5.10 shows the means and standard deviations for word length and task.

Analysis of error categories

Table 5.11 shows the errors categorized into transposition, blanks and intrusions. The table shows that under tapping conditions the word length effect is caused by differences in the number transposition. For the MIT lists there is also an effect of the number of blanks which goes in the same way as the effect for transposition. Under articulatory suppression, the difference in the number of transposition in Baddeley and MIT lists is removed. However, the effect of blanks is reversed in Baddeley materials but remains in the same direction for MIT materials. The number of intrusions does vary over the various conditions although the variations are considerably smaller than the variations in the other categories. However, it should be noted that for MIT materials the number of intrusions is doubled for long words compared to short words in the Tapping condition and this difference is absent under conditions of suppression.

Task	Vocabulary	Word length	Correct	Wrong	Blank	Intrusion	Total
Suppression	Baddeley	Short	346	85	141	28	600
		Long	358	89	118	35	600
		Total	704	174	259	63	1200
	MIT	Short	373	82	118	27	600
		Long	313	97	167	23	600
		Total	686	179	285	50	1200
Tapping	Baddeley	Short	422	50	107	21	600
		Long	391	87	106	16	600
		Total	813	137	213	37	1200
	MIT	Short	504	57	29	10	600
		Long	392	87	97	24	600
		Total	896	144	126	34	1200

Table 5.11: Individual responses by Task, Word length and Vocabulary. The Wrong category of error means a transposition.

Table 5.12 shows the three error categories broken down by whether the error occurred at a position where an item from part of a pair was in the stimulus or not. The table shows that under each condition the majority of the blanks occur for items which were not part of antonym pairs in that particular list.

In fact the correlation between the number of blanks in the recall and the number of antonym pairs in the stimulus for MIT materials is $r = -0.32, p < 0.001$ for short lists and $r = -0.30, p < 0.001$ for long lists and there are 131 stimulus pairs for short lists and 106 stimulus pairs for long lists in total. The correlation between the number of pairs in a list (of MIT materials) and the overall score is $r = 0.20, p < 0.001$ for short materials and $r = 0.12, p < 0.02$ for long materials.

Table 5.13 shows the size of the word length effect by two types of list. For lists solely made up of antonym pairs there is a small word length effect which is unaffected by suppression. For the lists which contain no pairs there is a negligible word length effect under tapping conditions which appears to be reversed under suppression conditions.

5.4.3 Discussion

The results presented in Table 5.2 suggest that the vocabulary sets do behave differently. The Baddeley vocabulary shows a word length effect in the Tapping condition which is abolished and possibly reversed in the Suppression condition. This replicates Baddeley, Thomson and Buchanan's (1975) result. However, the MIT vocabulary shows different behaviours across the two conditions. In the Tapping condition the words show the familiar word length effect but under Suppression there is still a significant large effect, although slightly diminished compared to the size of the effect under Tapping.

It might be suggested that the different behaviour of the word length effect between the two

Task	Word Length	Pair	Wrong	Blank	Intrusion	Total
Suppression	Short	no pair	76	109	25	210
		pair	6	9	2	17
		Total	82	118	27	227
	Long	no pair	83	151	22	256
		pair	14	16	1	31
		Total	97	167	23	287
Tapping	Short	no pair	49	22	9	80
		pair	8	7	1	16
		Total	57	29	10	96
	Long	no pair	78	89	23	190
		pair	9	8	1	18
		Total	87	97 24	208	

Table 5.12: Number of errors by category in Experiment XII by Task, Word length and Pair. For MIT lists, the errors have been distinguished by whether the item which should have been recalled was part of an antonym pair (in that particular list) or not.

	Suppression		Tapping	
Pairs	Short	Long	Short	Long
No pairs	2.3 (32)	2.7 (44)	3.0 (14)	3.0 (37)
All pairs	3.0 (11)	2.6 (12)	3.5 (10)	3.0 (7)

Table 5.13: Mean score for lists with no pairs and all pairs by task and word lenght (number is entries in each cell is reported in parentheses).

vocabularies was related to the size of the effect which is much bigger in the MIT vocabulary and can reasonably be attributed to differences in time-to-say between the two vocabularies. In particular the short words of the MIT vocabulary are much faster to say than the short words in the Baddeley vocabulary. It may be that suppression produces a fractional decrement of the word length effect and because the size of the effect is so large for the MIT words, although suppression has its expected effect, there is still a significant residue. This sort of explanation does not explain why the accuracy for long words is the same under Tapping but different under Suppression. If Suppression was behaving similarly in the two vocabularies then one would expect the long words in the two vocabularies to have similar accuracies, but this was not observed.

It might also be suggested that the subjects' performance of the secondary task varied. The analysis of the secondary task does hint at a possible effects of word length and task (the criterion for significance should probably be increased bearing in mind the multitude of *post hoc* comparisons) but there is no evidence for any interactions with vocabulary.

Baddeley's materials are intended to be semantically unrelated whereas the MIT materials are related. Inevitably interference could be offered as an explanation for the different behaviour of the MIT materials in that the lists of long words could cause more interference that the lists of short words. Articulatory suppression would have no effect on a semantic process so the apparent word length effect would remain with any articulatory component removed. The breakdown of errors showed that there was a category of errors which was affected by suppression (transpositions) and that there was a category of errors which was unaffected by suppression for MIT materials (blanks). If the blanks were being caused by interference then an interference explanation would stand.

Given that related items can interfere with each other, any pair of items, such as, *friendly* and *generous* might interfere with each other. However, it is more likely that antonyms will interfere with each other compared to non-antonymic items, simply because there is clear relationship between them. If that is the case then the greater the number of antonymic pairs the greater the interference. If blanks are being caused by interference then a positive relationship between number of antonymic pairs and blanks would be expected.

Table 5.12 shows the blank errors broken down by whether they occurred when an item was from an antonym pair or not (for MIT materials). The table clearly shows that the majority of the errors occur at positions in the lists which were occupied by items not accompanied by their antonyms. The correlation between number of pairs and blanks is negative and there are in fact more pairs in the short lists than in the long lists in the suppression condition. Therefore, the effect which is causing the blank errors does not seem to be simple interference caused by antonymic pairs. Not surprisingly the correlation between the number of pairs of items and the overall score for MIT lists is positive (although small) which shows that simple structure is helping memory. Although an interference explanation cannot be ruled out these results do not provide clear support for one.

However, there are inevitably other explanations which may depend on hidden lexemic variables like age of acquisition, imagability or concreteness. They may be correlated with word length and unaffected by suppression thus causing a spurious word length effect under suppression. Table 5.13 suggests that a lexemic explanation is inadequate. The same words or rather words drawn at random from the same pools are used to make up the two sorts of lists: ones with no antonym pairs and ones solely made up of antonym pairs. The means suggest that pairs of antonyms cause some sort of word length effect which is unaffected by suppression. However, lists without a pairwise structure show a negligible word length effect. If the two sorts of lists are composed of broadly similar lexemic items then any associated hidden variables must be operating in both types of list. There is an apparent difference in behaviour the word length effect cannot be attributed to correlated lexemic variables. Of course, the differences are very small and the number of measurements for each mean varies so the results can be suggestive at best. Only further experimentation can resolve these issues.

Even if the explanation for the results remains uncertain the point to be emphasised is that lists of words with some semantic pairings (antonymic) do behave differently from unstructured lists. Clearly, the texts used in an MIT do have structure although of a more complex sort than in these last two experiments. Therefore the results suggest that the lack of an interaction between articulatory suppression and word length in Experiment X may be due to the structured nature of the texts and not due to subjects' inadequate performance of the secondary task or of word frequency. The simplest explanation of the results is that with materials which are semantically structured genuine articulatory word length effects arise which are not susceptible to articulatory suppression.

5.5 General discussion

Experiment X showed that although robust word length effects were found in the reading time and recall data in an MIT they did not interact with articulatory suppression. This lack of interaction threw some doubt on the interpretation of word length effects as genuine articulatory phenomena of the sort described by Baddeley, Thomson and Buchanan (1975). Articulatory suppression seemed to have a very similar effect to the Tapping control which implied that it was only imposing an attentional load on the subjects. However, there were instances where Suppression had an effect which was different from both Normal and Tapping conditions. Suppression therefore appeared to have had a subtle effect. There were two possible reasons why suppression had had such a small effect. First the subjects might not have been rehearsing the irrelevant item conscientiously or second, there may have been unknown effects due to the particular materials which correlated with word length.

Experiment XI compared the materials from Experiment X to materials used in Baddeley, Thomson and Buchanan (1975). The analysis showed that particular cohorts behaved like the Baddeley materials and others showed a word length effect which was resistant to articulatory suppression. The cohorts which showed the resistant word length effects were cohorts which were composed of antonym pairs and therefore had more semantic structure than the other cohorts. Because frequency and suppression rate had not been properly controlled these explanations could not be ruled out.

Experiment XII was very similar to Experiment XI except that a new set of materials were compared to Baddeley's. In these materials frequency was controlled and while the task was being carried out subjects' behaviour was strictly monitored. The same result was found in that the MIT materials still showed a word length effect under articulatory suppression.

Experiments which investigate word length and suppression (*e.g.*, Baddeley, Thomson & Buchanan, 1975; Besner & Davelaar, 1982; Coltheart, Avons & Trollope, 1990) generally find that the word length effect is reduced if not abolished by suppression. However, as Besner (1987) points out, suppression does not necessarily have an effect on the phonological codes mediating lexical access. Besner and Davelaar (1982) showed that words with any entry in the phonological lexicon (*e.g.*, BRANE) were better recalled than words which had no entry (*e.g.*, SLINT). This difference persisted under articulatory suppression which they interpreted as evidence for a suppression resistant phonological code. It is unlikely that the word length effects in the MIT are due to these suppression resistant codes because Besner and Davelaar (1982, Experiment II) did find a significant interaction between word length and suppression even though there remained a significant lexicality effect.

The main finding of the experiments reported here, that there is a suppression resistant word length effect does not seem to fit very well with previous results (*e.g.*, Baddeley, Thomson & Buchanan, 1975). However, Bishop and Robson (1989) report some relevant results with speechless adults. They reasoned that individuals who cannot generate articulatory codes should be unable to translate written stimuli into phonological representations which would mean that such individuals would not show a word length effect or a phonological similarity effect. Essentially such individuals should show the same behaviour as subjects in an experiment using articulatory suppression.

Previous work which had investigated speechless adults had used subjects with acquired anarthia (Nebes, 1975; Baddeley & Wilson, 1985; Vallar & Cappa, 1987). These studies showed a normal pattern, such as, word length effects and phonemic confusions. Baddeley (1986, p. 177) proposed that the results might be explained by assuming that speechless adults retained the ability to generate motor programs for speech but were inhibited from enacting the programs at a peripheral level. Bishop and Robson (1989) reasoned that congenitally speechless individuals would never have had the opportunity to learn even the motor programs necessary for speech and would therefore be better subjects in an investigation of speechlessness and short term memory.

They investigated congenitally anarthic (never having been able to speak), dysarthic (laboured and often unintelligible speech) and matched normal controls. An eye pointing task was used in which a "menu" of pictures was displayed and subjects selected items from the menu by gazing at them. This procedure was used to estimate memory span. Subjects were shown a sequence of pictures and asked to recall them in sequence by eye-pointing. The pictures corresponded to words drawn from one of three sets: a control set (*e.g.*, bath, pig, leaf), a long-named set (*e.g.*, elephant, aeroplane, kangaroo) and a phonologically similar set (*e.g.*, man, van, pan). Memory span was estimated for each set by showing subjects lists of increasing length until a mistake was made and the memory span was therefore the greatest number of correctly recalled items. The results showed that the control items were better remembered than the phonologically similar and polysyllabic lists across all groups of subjects. They concluded that the phonological similarity and word length effects in short-term memory did not depend on subvocal articulatory gestures. In addition, they suggested that rehearsal is possible even for people who have never spoken if it is accepted that the word length effect arises from totally non-articulatory subvocal rehearsal. They carried out a second experiment using a rhyme judgement task and concluded that articulatory coding was unnecessary for the task.

As an explanation for their result's apparent disagreement with the conventional results from the short-term memory literature they suggested that there is an abstract phonological representation which is used by a speech-motor-program generator but does not itself contain an articulatory specification. Although they cannot explain why suppression does disrupt the word length effect and phonemic similarity effect in normal subjects they propose that the disruption is not due to interference with articulatory coding. Unfortunately, they did not test their initial assumption that anarthic subjects behaved the same way as normal subjects under articulatory suppression conditions. Therefore their results do not generalise as well as they intended.

An investigation by La Pointe and Engle (1990) of the difference between simple and complex word span measures contains an experiment very close to the one which Bishop and Robson's (1989) investigation is missing. La Pointe and Engle (1990) were interested in the difference between simple word span-measures which did not predict reading ability (Perfetti & Lesgold, 1977) and reading span-measures which did (Daneman & Carpenter, 1980). They reasoned that it was crucial to investigate how the two measures responded to different variables. Simple word-span tasks were well known to exhibit articulatory phenomena, like the word length effect, so they concluded that it would be interesting to find out if a reading span-measure also produced a word length effect.

La Pointe and Engle's reading span-task was very similar to the task used by Daneman and Carpenter (1980). Subjects were required to read aloud a set of sentences, each of which was followed by a word (the test item) not in the sentence. When the set had been read the subject was required to write down the test words in any order. When the subject had completed three trials with a particular number of sentences in the set, the number was increased and another set of three trials initiated. After three trials of five sentence sets had been completed the reading span task was completed. The word span test was very similar except that no sentences were presented, only the test items at a rate of 1 per second. Also the test was not complete until a set of seven (as opposed to five in the reading span task) words had been used.

The results showed a word length effect for both measures across a variety of different scoring methods which they concluded demonstrated that the two tasks measured similar aspects of processing. They further showed that the word length effect in the complex (reading) task was not affected if the subjects were required to verify arithmetic statements instead of reading (which they called an operation span task). In their third experiment they tested whether the word length effects they had observed were affected by a suppression task in which subjects were required to repeat "abcabc..." continuously during presentation. They had supposed that the reason for the divergent predictive power of the two measures might lie in the use of different coding, one which might be sensitive to suppression, the other not (Besner & Davelaar, 1982). The results showed that although the the measures were diminished in comparison to previous results the word length effect remained. In order to explain this effect, when compared with Baddeley, Thomson & Buchanan's (1975) result, they again postulated different codes. In their tasks different words were used on every list whereas in Baddeley *et al.*'s task there was a considerable overlap and they proposed that using different words engendered a deeper code which was unaffected by suppression.

Their fourth experiment used two sets of words: one fixed in size and one unlimited (very large) in size. Again, both tasks were used (word span and operation span) and suppression and no suppression conditions were imposed. Their results showed that for a fixed pool of words there was a word length effect which was abolished by suppression and that this interaction between suppression and word length appeared in both tasks. For an unlimited pool of words, however, the results showed a word length effect across both tasks irrespective of suppression. In their fifth experiment they investigated a confounding factor of concreteness and whether serial *vs.* free recall made any difference while repeating their fourth experiment. The same pattern of results was found irrespective of recall procedure.

These results clearly relate to the results reported in this chapter in that a word length effect was found which was unaffected by articulatory suppression. Of course a different measure was used and a different manipulation was found to affect the word-length suppression interaction, namely pool size. Even so, the explanation that they propose may help explain the results reported here. Essentially, they view the articulatory loop, not as an "inherent structural aspect of human cognition" but as a "combination of a code that decays over time and a time-limited rehearsal process". When this is accepted then it is supposedly possible to envision a whole continuum of codes, of which phonological and articulatory ones are just two examples.

They cite Reisberg, Rappaport and O'Shaughnessy (1984) who were able to teach subjects a finger coding strategy for use in memory tasks and who proposed that subjects would use any coding strategy which was useful in a memory task. Therefore, they propose that the large pool used in their experiments allowed a different coding strategy to be used which did exhibit a word length effect but was unaffected by suppression.

Perhaps, a similar sort of code was developed by subjects in Experiment X because of the extra semantic structure available in the materials, even although they were drawn from a fixed pool. They do not advance an explanation for the difference between the coding which depends upon the two sizes of pool but when their results are compared to the ones here perhaps a general explanation can be arrived at. In the experiments reported here, the presence of antonyms in a supra span test seem to induce a suppression resistant word length effect and in the La Pointe and Engle (1990) study span (four different measures) was affected by word length and the interaction with suppression was affected by the size of the pool from which the words were drawn.

The commonality between the manipulations is semantic interference. By increasing the size of the pool of words from which the test items are drawn the repetition of words across lists was reduced to zero. Therefore the difference between the conditions could be expressed as limited *vs.* zero inter-list interference. In the antonym experiments reported here the contrast could be stated as the difference between two levels of intra-list interference if it is assumed that antonyms interfere more than unrelated pairs. Although the contrasts do not allow a unification of the results they do point to the involvement of semantic information in an effect which is meant to be asemantic. Therefore the possibility arises that a semantic contrast forces the use of some coding scheme which produces a suppression-resistant articulatory code which is word length related.

The question of what exactly is happening in working memory to explain these effects is still unanswered. Unfortunately the status of working memory itself is unclear. Baddeley (1986) has described his unitary conception, Monsell (1984) his distributed view and both Bower (1975) and Anderson (1983) distinguish between working memory and short term memory by supposing that working memory's contents are not available to conscious examination. The issue of word length effects is also difficult to assess because Baddeley (1986) attributes them to an input buffer whereas Monsell (1987) attributes them to an output buffer and Fitzgerald, Tattersall and Broadbent (1988) propose an input-output buffer as well. Further difficulties arise from the neuropsychological literature and some are described in Vallar and Shallice (1990). In particular, Howard and Franklin (1990) cast doubt on the use of rehearsal in short-term recall. The explanation of the effects observed here may be any one of the ones mentioned above and the solution lies in further experimental and theoretical work.

The discovery of word-length effects in the MIT was particularly interesting because they implied that a well investigated area of working memory might be involved in the task and that working memory theory might be integrated or exploited in a theory of the construction of the representations of individuals. However, the experiments reported in this chapter make it clear that the working memory theory is not complete and that the effects on which Baddeley's (1986) theory are based are more complicated than at first thought. Although the word-length effects were not abolished in Experiment X this is not evidence against their interpretation as verbal memory effects. In addition these experiments form part of a body of converging evidence which shows that the model of the articulatory loop must involve some semantic apparatus (*e.g.*, Hulme, Maughan & Brown, in press).

Chapter 6

General discussion and final remarks

Summary

People use a particular strategy to restore parallelism in a text because parallelism is a key feature of memory which underlies text processing. This effect of parallelism has been observed in at least two different paradigms and may represent a link between implicit and explicit memory systems. A subsidiary finding of this investigation throws doubt upon the interpretation of effects which were originally taken to indicate the use of phonological memory.

6.1 Summary of results

THE RESULTS reported in Chapter 2 show that varying the order of reference in simple texts has a considerable effect compared to the results collected using predictable texts in Stenning, Shepherd and Levy (1988). An asymmetry between the two individuals appears in the recall data which is also found in the reading time data. The primary individual is usually associated with the first recalled individual and exhibits the SOE when there is a switch of reference to it in the reading time data. Assuming that studying switches of reference explores the links between different stores these properties seem to imply that the primary individual is held in a long term store whereas the secondary individual has a much less permanent existence. The question arises why an asymmetry should appear and Stenning's (1991) analysis shows that the asymmetry is a method of producing order out of the apparent chaotic patterns of reference so that the simple relationship between surface order and semantics is restored. This restoration is essentially a restoration of parallelism in the text and is possibly a primitive representational mechanism.

The next chapter, Chapter 3 contains a description of a series of experiments which were designed to investigate the possibility that if parallelism was a property generally exploited in language processing then it should be observable in a task other than the MIT. The experiments were all aimed at studying parallelism (both grammatical and sequential) in pronoun comprehension. Both comprehension and resolution were studied, and although the effects appeared weak, there was evidence that parallelism was being exploited in comprehension. Further, the results suggested that parallelism was not simply a sentence level phenomenon as had been supposed in the literature but operated over several sentences as it had done in the MIT. Subsidiary results from the experiments demonstrated that animacy and sentence structure affected the use of parallelism. They also showed that there was an order of mention effect which is closely related to a subject assignment strategy.

The fourth chapter returned to some issues which had been revealed by the experiments in Chapter 2. The regression models had shown that description length effects were present in the reading time data which had been interpreted as articulatory rehearsal using Baddeley's (1986) theory of working memory. The most robust effect was a negative effect which showed that reading time decreased in proportion to the length of the secondary individual's description when there was a switch of reference to the primary individual. In order to corroborate this effect a simple MIT was run and subjects were asked to externalise their rehearsal. When the data was analysed it was clear that subjects were abbreviating their rehearsal of the primary individual when there was a switch of reference to the primary individual which was interpreted as evidence for the negative effect described above. The data also revealed individual differences between subjects which were attributed to differences in the interpretation of the instructions.

Although the word length effects in the regression models were supported by the results from the overt rehearsal experiment there was still some doubt about their cause. Therefore an MIT was run with articulatory suppression which is known to remove word length effects in simple word list experiments. The data, reported in Chapter 5, showed that there was a slight effect of suppression in the recall data on confusions of recall order when there had been a change of status. However, the reading time data and the recall data showed a significant word length effect which was unaffected by suppression. As a way of explaining this result a series of small list recall experiments were run which showed that frequency alone could not explain these results and that there appeared to be semantic or long term memory effects present. The effects seemed to cause some sort of articulatory coding which was not susceptible to articulatory suppression.

6.2 Conclusions

The introduction discussed the reason why studying patterns of reference would reveal aspects of memory not amenable to study using the more common list learning paradigms or rich naturalistic prose comprehension/memory tasks. It was hoped that by studying such patterns and developing an understanding of their origin in terms of memory architecture an account of why structure in text is the way it is would be adduced. Furthermore it was supposed that working memory was inevitably involved in the processing of text and Baddeley's (1986) theory was the most developed so it would naturally be part of the explanation.

Clearly the experimental work reported here has advanced the understanding of these issues. First, by building on Stenning and Levy's (1988) model of human binding an explanation of the observed asymmetries in the MIT has been developed. This explanation depends on a mapping between surface order and semantics being an inherent part of the representational machinery used in text comprehension and memory. Because this machinery is fixed an extra layer of labelling must be imposed to allow it to function thus causing the observed asymmetries. These asymmetries are related to the concept of focus (Sanford & Garrod, 1982) and offer a possible explanation of such findings in richer more naturalistic texts.

The conclusion that human memory does store sequence information is paradoxical at first glance. Studies of sentence memory using direct measures (such as, free recall and recognition) have shown that surface features of sentence memory are quickly forgotten (Begg & Wickelgren, 1974; Sachs, 1974) but other work using indirect measures suggests that surface information can have long-lasting effects (Kolers, 1975; Masson & Sala, 1978). Although the type of surface information referred to varies it is clear that implicit memory (Schacter, 1987) does retain different more superficial information compared to explicit memory. Therefore, there may not in fact be a paradox and the proposal that the explicit recall of a pattern of binding inferred from a lower-level representation using surface information may in fact be an inference from implicit to explicit memory.

Second, having identified parallelism in the MIT, parallelism has been found in pronoun comprehension. The literature contained work which had claimed to have found parallelism but few studies had really tackled the problem adequately. Therefore the work here is an advance because the effect has been isolated and furthermore its interactions with other properties of language have been studied. Of course, the main thrust of this work on pronouns is in relation to the work on the MIT but it clearly stands on its own as well.

The work on pronoun resolution provides support for the hypothesis that serial-order information is encoded and used in the resolution process. However, there are clearly many potential strategies operating in these experiments and inevitably some of them are confounded. Therefore further experimentation is needed to separate them which will depend mainly on which grammatical constructions can be found. The issue of the difference between subject and object pronouns is particularly interesting and might benefit from an investigation using an eyetracker. This tool would be particularly good for examining the processes before context information became available. The point has been made that simple strategies operate when more complicated ones depending on pragmatics cannot but it may also be the case that they operate all the time but that they are simply overridden by later processes, so increasing the temporal resolution of the analytic tool makes it possible to study these simple strategies irrespective of the effects of general knowledge.

The work in the construction of representation of individuals clearly establishes an asymmetry of processing as an emergent property of the representational machinery which depends on inference. Several questions follow from this. Just as order effects can be found in pronoun comprehension are these asymmetries also likely to be found? There is an immediate parallel with work on focus and the question is really, do the two effects relate. One possible way of examining this question might be to introduce pronouns in the MIT and see if they were more likely to be assigned to the primary individual, for example. The MIT might be further extended to separate the two effects of property attribution and switching reference by presenting each sentence of an MIT as two clauses: subject followed by predicate. Lastly, given the evidence of serial-order encoding can this information be incorporated within Stenning and Levy's (1988) model of binding? Unfortunately there do seem to be difficulties with encoding order of mention as with quantificational facts although it is conceivable that content could bind such propositions. After all "fat" followed by "thin" may be a slimming idea whereas as "thin" followed by "fat" may be a gluttony idea.

The third main conclusion may appear a negative one but is in fact positive. Baddeley's (1986) theory of working memory has enjoyed success in accounting for the phenomena described in Baddeley (1986) but the work here has shown that the phenomena are in fact more subtle and therefore the theory is in need of some revision. Other work has been suggesting a similar view and indeed some of Baddeley's latest work (Papagno, Valentine & Baddeley, 1991) on language learning is clearly addressing these new issues. The work also shows that the interpretation of the word length effects is still valid and clearly some sort of verbal memory is being exploited alongside a richer semantic one. Given the order retaining properties of verbal memory this is no surprise because order is clearly implicated in the representational apparatus used in attribute binding.

There may appear to be a contradiction between the methodology used to investigate the attribute binding problem and the methodology used to investigate working memory. Both approaches, so far, have advocated the use of tasks which strain subjects' abilities beyond their normal limits because a complex system's break down may well be very informative (Dell, 1968). As mentioned above, this approach may have been slightly misleading research in working memory because subjects may have relied on rarely used strategies to cope with the supra-span tasks so often used. Given that subjects inevitably use unusual strategies in abnormal conditions why should the MIT results be reliable? Clearly, this is a serious issue which can be addressed by considering the breadth of the evidence which supports the general conclusions drawn from the MIT work. Basically, asymmetries are common in language processing and may relate to focus phenomena; switching reference unpredictably takes time which fits with results from the focussed attention literature and parallelism phenomena have been found in pronoun comprehension.

The MIT may appear at first sight an artificial task but it has revealed an aspect of processing which might otherwise have remained obscure had more complicated materials been used. This aspect of processing has been generalised to pronoun comprehension which gives it more weight. This analysis of reference ordering in text has gone some way to providing an explanation of why we structure a text the way we do.

6.3 Future work

There are several ways in which issues investigated here could be pursued. As mentioned above an eyetracker might be useful in studying parallelism in pronoun comprehension, pronouns would be an interesting addition to the MIT as an aid to investigating focus and Stenning and Levy's (1988) model could be extended to incorporate order information. There are several further lines of investigation which could be pursued: working memory, content, memory updating and modality differences.

One of the motivations for postulating different stores in memory was the different effects of semantics in different tasks. The investigation of working memory reported here, suggests that this distinction may no longer be so clear. Obviously there are still many possible explanations for the effects described here which relate to lexical features such as concreteness and age of acquisition. However, there is a growing body of other experimental findings which cast doubt on Baddeley's simple explanations (*e.g.*, Brooks & Watkins, 1990; Tehan & Humphreys, 1988). Research, subsequent to the research here could fruitfully follow an exploration of the interference effects which might be operating and would relate to La Pointe and Engle's (1990) study. However, there is also a growing body of research which shows that different measures and different tasks (Hulme, Maughan & Brown, in press) cause people to use markedly different strategies which produce different results and that the apparent simplicity and uniformity only comes from averaging over large populations. Therefore future investigations must address the issue of which tasks most relate to people's everyday use of working memory and one of the lessons of this research in relation to the MIT is that structure is an everyday occurrence which must be addressed.

Given the emphasis on content and general knowledge in the introduction there is clearly a need to empirically investigate its deployment in the binding problem. Nelson (1988) reports an attempt to investigate subjects' knowledge of stereotypes in relation to the MIT and finds that there are weak effects. Unfortunately quantifying content is a very awkward problem and restricting the problem to stereotypes does indeed reduce the scope of the problem but subtleties still remain. For example, some stereotypes are very unusual and others are very common so such distinctions must be taken into account. However, Nelson (1988) did demonstrate that subjects' introspection about stereotypes is structured so that measure may be a useful way to investigate the relation of content and binding.

Memory updating is the act of modifying the current status of a representation to accommodate new input. This is clearly what subjects are doing when they recall an individual's description when they switch reference to the primary individual in an MIT text. This process also involves some attentional control and therefore the central executive (Morris & Jones, 1990) which has remained relatively obscure in working memory theory. By developing a model of this process perhaps a better understanding of the processes used to orchestrate updating could be achieved. Such an understanding may inform ideas about memory load.

Many investigations postulate memory loads at various points in a task. For example, Ehrlich and Johnson-Laird (1982) suppose that information which cannot be immediately incorporated into a mental model of a spatial description will impose a memory load and find evidence for this from elevated reading times for the corresponding sentences. Similarly, Wanner and Maratsos (1978) propose that a dislocated constituent imposes a memory load until its "hole" is encountered. Although there may be a correlation between the incidence of a hypothesized memory load and an unusually long reading time, the memory load does not actually explain the increase in reading time. After all, we don't slow down through our lifetime as we accrue new information. Perhaps some sort of "swapping" process is taking place which swaps information or to-be-remembered facts between different stores. If this were taking place then perhaps the dislocated constituent is being stored in some other buffer from which it must be recalled as possible holes are encountered.

By studying such "swapping" processes perhaps a new method of investigating memory capacity might be developed. On modern computers, when the machine's easily accessed immediate memory is exceeded the machine can read and write information to a hard disk very quickly and thus accommodate an overflow. Although the process is very fast the machine's performance may well slow down. Therefore, if human memory swapping is at all related then the slowing down of a process may indicate swapping and therefore memory overflow. By studying unpredictable switches of reference then, swapping may be induced (as is suggested by the model described in Chapter 4) and capacity characteristics could conceivably be inferred. Although this view is clearly one inspired by modern computers it may still be a useful descriptive tool in understanding human abilities.

Jakimik and Glenberg (1990) investigated modality differences in pronoun comprehension using Glenberg and Swanson's (1986) temporal distinctiveness theory. This theory proposes that there is a difference in coding between visual and auditory modalities such that order information is more accurately represented if presented auditorily. Glenberg and Fernandez (1988) found that memory for order of occurrence was better for items heard rather than seen. Jakimik and Glenberg (1990) were concerned to explain why temporal anaphora (for example, "former/latter", "first/second") are difficult to process in written language. Temporal distinctiveness theory predicted that these anaphors would be easier to understand when presented auditorily which is what their results demonstrated from which they conclude that some forms of anaphora are sensitive to surface order. Therefore the role of order in the MIT and attribute binding might be investigated by exploiting modality differences.

This distinction between modalities may explain why the ancient logographers became sensitized to ordering phenomena in text. After all, their speeches were intended to be memorised for oral delivery.

> It is best to do things systematically, since we are only human, and disorder is our worst enemy.
> HESIOD, *Works and Days*, 471

References

Anderson, J. R. (1972) FRAN: A simulation model of free recall. In G. H. Bower, ed., *The Psychology of Learning and Motivation*, Vol. 5. New York: Academic Press.

Anderson, J. R. (1976) *Language, Memory and Thought.* Hillsdale, N.J.: Lawrence Erlbaum Associates.

Anderson, J. R. and G. H. Bower (1973) *Human Associative Memory.* Washington D.C.: Hemisphere.

Anderson, J. R. (1983) *The Architecture of Cognition.* Cambridge, Mass.: Harvard University Press.

Aristotle (1991) *The Art of Rhetoric.* Harmondsworth, Middlesex: Penguin. Translated by Lawson-Tancred.

Atkinson, R. C. and M. R. Raugh (1975) An application of the mnemonic keyword method to the acquisition of a Russian vocabulary. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology* **104**, 126–133.

Baddeley, A. D. (1981) Reading and working memory. *Bulletin of the British Psychological Society* **35**, 414–417.

Baddeley, A. D., M. Eldridge and V. J. Lewis (1981) The role of subvocalization in reading. *Quarterly Journal of Experimental Psychology* **33**, 439–454.

Baddeley, A. D. and G. Hitch (1974) Working Memory. In G. H. Bower, ed., *The Psychology of Learning and Motivation*, Vol. 8, pp. 47–90. New York: Academic Press.

Baddeley, A. D. and V. J. Lewis (1981) Inner active processes in reading: the inner voice, the inner ear, and the inner eye. In A. M. Lesgold and C. A. Perfetti, eds., *Interactive processes in reading.* Hillsdale, N.J.: Lawrence Erlbaum Associates.

Baddeley, A. D., N. Thomson and M. Buchanan (1975) Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior* **14**, 575–589.

Baddeley, A. D. and B. Wilson (1985) Phonological coding and short-term memory in patients without speech. *Journal of Memory and Language* **24**, 490–502.

Baddeley, A. (1986) *Working Memory.* Oxford: Oxford University Press.

Baddeley, A., G. Vallar and B. Wilson (1987) Sentence Comprehension and Phonological Memory: Some Neuropsychological Evidence. Chap. 23 in *Attention and Performance*, Vol. XII: *The psychology of reading.* Hove: Lawrence Erlbaum Associates.

Baddeley, A. D. (1976) *The Psychology of Memory.* New York: Basic Books.

Bairaktaris, D. (1990) A model of Auto-Associative Memory that stores and retrieves, Successfully, data regardless of their Orthogonality, Randomness or Size. In *Hawaii International Conference on System Sciences '90.*

Bartlett, F. C. (1932) *Remembering: a study in experimental and social psychology.* Cambridge: Cambridge University Press.

Begg, I. and W. A. Wickelgren (1974) Retention functions for syntactic and lexical vs. semantic information in sentence recognition memory. *Memory and Cognition* **2**, 353–359.

Berwick, R. C. and A. Weinberg (1984) *The Grammatical Basis of Linguistic Performance: Language Use and Acquisition.* Cambridge, Mass.: MIT Press.

Besner, D. and E. Davelaar (1982) Basic Processes in Reading: Two Phonological Codes. *Canadian Journal of Psychology* **36**, 701.

Besner, D. (1987) Phonology, Lexical Access in Reading, and Articulatory Suppression: A Critical Review. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology* **39**, 467–478.

Bishop, D. V. M. and J. Robson (1989) Unimpaired short-term memory and rhyme judgemet in congenitally speechless individuals: Implications for the notion of "articulatory coding". *Quarterly Journal of Experimental Psychology: Human Experimental Psychology* **41A**, 123–140.

Bower, G., J. B. Black and T. J. Turner (1979) Scripts in Memory for Text. *Cognitive Psychology* **11**, 177–220.

Bower, G. H. (1973) How to... Uh... remember! *Psychology Today* **7**, 63–70.

Bower, G. H. (1975) Cognitive Psychology: An introduction. In W. K. Estes, ed., *Handbook of learning and cognitive processes*, Vol. 1. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Bower, G. H. and M. C. Clark (1969) Narrative stories as mediators for serial learning. *Psychonomic Science* **14**, 181–182.

Bransford, J. D., J. R. Barclay and J. J. Franks (1972) Sentence memory: a constructive versus interpretive approach. *Cognitive Psychology* **3**, 193–209.

Bransford, J. D. and M. Johnson (1972) Contextual prerequisites for understanding:some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior* **11**, 717–726.

Brooks III, J. O. and M. J. Watkins (1990) Further Evidence of the Intricacy of Memory Span.

Journal of Experimental Psychology (Learning, Memory and Cognition) **16**, 1134–1141.

Caramazza, A., E. Grober, C. Garvey and J. Yates (1977) Comprehension of anaphoric pronouns. *Journal of Verbal Learning and Verbal Behavior* **16**, 601–609.

Caramazza, A. and S. Gupta (1979) The roles of topicalization, parallel function and verb semantics in the interpretation of pronouns. *Linguistics* **17**, 497–518.

Carroll, J. M. and T. G. Bever (1976) Sentence comprehension: a case study in the relation of knowledge and perception. In E. C. Caterette and M. P. Friedman, eds., *Handbook of Perception*, Vol. 7: *Language and Speech*, pp. 209–344. New York: Academic Press.

Chafe, W. (1976) Givenness, contrastiveness, definiteness, subjects, topics and points of view. In C. Li, ed., *Subject and Topic*, pp. 25–56. New York: Academic Press.

Chafe, W. L. (1972) Discourse structure and human knowledge. In J. B. Carroll and R. O. Freedle, eds., *Language comprehension and the acquisition of knowledge.* Washington, DC: Winston.

Chambers, S. M. (1979) Letter and order information in lexical access. *Journal of Verbal Learning and Verbal Behavior* **18**, 225–241.

Charness, N. (1979) Components of skill in bridge. *Canadian Journal of Psychology* **33**, 1–50.

Chomsky, N. and H. Lasnik (1977) Filters and Control. *Linguistic Inquiry* **8**, 425–504.

Cirilo, R. K. (1981) Referential coherence and text structure in story comprehension. *Journal of Verbal Learning and Verbal Behavior* **20**, 358–367.

Clark, H. H. (1969) Linguistic processes in deductive reasoning. *Psychological Review* **76**, 387–404.

Clark, H. H. and E. V. Clark (1977) *Psychology and Language: An Introduction to Psycholinguistics.* New York: HCB.

Clark, H. H. and C. J. Sengul (1979) In search of referents for nouns and pronouns. *Memory and Cognition* **7**, 35–41.

Cohen, J. (1983) *Applied Multiple Regression.* Hove: Lawrence Erlbaum Associates.

Collins, A. M. and E. F. Loftus (1975) A Spreading Activation Theory of Semantic Processing. *Psychological Review* **82**, 407–428.

Collins, A. M. and M. R. Quillian (1969) Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior* **8**, 240–247.

Collins, A. M. and M. R. Quillian (1970) Does category size affect categorisation time? *Journal of Verbal Learning and Verbal Behavior* **9**, 432–438.

Coltheart, M. (1981) The MRC psycholinguistic database. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology* **33A**, 497–505.

Coltheart, V., S. E. Avons and J. Trollope (1990) Articulatory suppression and phonological codes in reading for meaning. *Quarterly Journal of Experimental Psychology* **42A**, 375–399.

Conrad, C. (1972) Cognitive economy in semantic memory. *Journal of Experimental Psychology* **92**, 149–154.

Conrad, R. (1964) Acoustic confusion in immediate memory. *British Journal of Psychology* **55**, 75–84.

Conrad, R. (1970) Short-term memory processes in the deaf. *British Journal of Psychology* **61**, 179–195.

Conrad, R. and A. J. Hull (1964) Information, acoustic confusion and memory span. *British Journal of Psychology* **55**, 429–432.

Cowan, J. R. (1980) The significance of parallel function in the assignment of anaphora. In J. Kreiman and A. E. Ojeda, eds., *Papers from the Parasession on Pronouns and Anaphora at the Seventeenth Regional Meeting of the Chicago Linguistics Society*. Chicago:

Craik, F. I. M. and R. S. Lockart (1972) Levels of processing: A framework for memory reserach. *Journal of Verbal Learning and Verbal Behavior* **11**, 671–684.

Crawley, R. A., R. J. Stevenson and D. Kleinman (1990) The Use of Heuristic Strategies in the Comprehension of Pronouns. *Journal of Psycholinguisitc Research* **19**, 245–264.

Daneman, M. and P. A. Carpenter (1980) Individual Differences in Workign Memory in Reading. *Journal of Verbal Learning and Verbal Behavior* **19**, 450–466.

De Groot, A. D. (1966) Perception and memory versus thought: Some old ideas and recent findings. In B. Kleinmuntz, ed., *Problem Solving*. New York: John Wiley and Sons.

De Soto, C. B., M. London and S. Handel (1965) Social reasoning and spatial paralogic. *Journal of Personality and Social Psychology* **2**, 513–521.

Dell, G. S. (1986) A spreading activation theory of retrieval in sentence production. *Psychological Review* **93**, 283–321.

Dixon, W. J. and et al., eds. (1983) *BMDP Statistical Software*. University of California Press.

Dowty, D. (1991) Thematic Proto-Roles. *Language* **67**, 547–619.

Draper, N. R. and H. Smith (1981) *Applied Regression Analysis*, 2nd ed. Chichester: John Wiley and Sons.

Egan, D. E. and B. J. Schwartz (1979) Chunking in recall of symbolic drawings. *Memory and Cognition* **7**, 149–158.

Ehrlich, K. and P. N. Johnson-Laird (1982) Spatial descriptions and referential continuity. *Journal of Verbal Learning and Verbal Behavior* **21**, 296–306.

Eisenstadt, M. and Y. Kareev (1975) Aspects of human problem solving: The use of internal representations. In D. A. Norman and D. E. Rumelhart, eds., *Explorations in cognition.* San Fransisco: W. H. Freeman.

Engle, R. W. and L. Bukstel (1978) Memory processes among bridge players of differing expertise. *American Journal of Psychology* **91**, 673–690.

Ericson, K. A. and H. A. Simon (1984) *Protocol analysis: verbal reports as data.* Cambridge, Mass.: MIT Press.

Ericsson, K. A. and P. G. Polson (1988) A Cognitive Analysis of Exceptional Memory for Restaurant Orders. In M. T. H. Chi, R. Glaser and M. J. Farr, eds., *The Nature of Expertise*, pp. 23–70. LEA.

Eysenck, M. W. and M. C. Eysenck (1980) Effects of processing depth, ditinctiveness, and word frequency on retention. *British Journal of Psychology* **71**, 263–274.

Fischler, I., D. Rundus and R. C. Atkinson (1970) Effects of overt procedures on free recall. *Psychonomic Science* **19**, 249–250.

FitzGerald, P., A. Tattersall and D. Broadbent (1988) Separating central mechanisms by POCs: Evidence for an input-output buffer. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology* **40A**, 109–134.

Francis, W. N. and H. Kucera (1982) *Frequency Analysis of English Usage: Lexicon and Grammar.* Boston, Mass.: Houghton Mifflin.

Frazier, L., L. Taft, T. Roeper, C. Clifton and K. Ehrlich (1984) Parallel structure: A source of facilitation in sentence comprehension. *Memory and Cognition* **12**, 421–430.

Fredriksen, J. R. (1981) Understanding anaphora: Rules used by readers in assigning pronominal referents. *Discourse Processes* **4**, 323–347.

Garnham, A., J. Oakhill and P. N. Johnson-Laird (1982) Referential continuity and the coherence of discourse. *Cognition* **11**, 29–46.

Garrod, S. C. and A. J. Sanford (1982) The mental representation of discourse in a focussed memory system: implications for the interpretation of anaphoric noun phrases. *Journal of Semantics* **1**, 21–42.

Garvey, C. and A. Caramazza (1974) Implicit causality in verbs. *Linguistci Theory* **5**, 459–464.

Garvey, C., A. Caramazza and J. Yates (1976) Factors influencing assignment of pronoun antecedents. *Cognition* **3**, 227–243.

Gernsbacher, M. A., D. J. Hargreaves and M. Beeman (1989) Building and Acessing Clausal Representations: The Advantage of First Mention versus the Advantage of Clause Recency. *Journal of Memory and Language* **28**, 735–755.

Gernsbacher, M. A. and D. J. Hargreaves (1988) Accessing Sentence Participants: The Advantage of First Mention. *Journal of Memory and Language* **27**, 699–717.

Glass, A. L. and K. J. Holyoak (1975) Alternative Conceptions of Semantic Memory. *Cognition* **3**, 313–339.

Glenberg, A. M. and A. Fernandez (1988) Evidence for auditory temporal distinctiveness: Modality effects in order and frequency judgements. *Journal of Experimental Psychology (Learning, Memory and Cognition)* **14**, 728–739.

Glenberg, A. M. and N. C. Swanson (1986) A temporal distinctiveness theory of recency and modality effects. *Journal of Experimental Psychology (Learning, Memory and Cognition)* **12**, 3–15.

Goodluck, H. and S. Tavakolian (1982) Competence and processing in children's grammar of relative clauses. *Cognition* **11**, 1–27.

Gotoff, H. C. (1979) *Cicero's Elegant Style: An Analysis of the* Pro Archia. Urbana, Ill.: University of Illinois Press.

Greenberg, J. H. (1963) Some universals of grammar with particular reference to the order of meaningful elements. In J. H. Greenberg, ed., *Universals of language*, pp. 58–90. Cambridge, Mass.: MIT Press.

Greeno, J. G. and D. L. Noreen (1974) Time to read semantically related sentences. *Memory and Cognition* **2**, 117–120.

Grober, E., W. Beardsley and A. Caramazza (1978) Parallel function strategy in pronoun assignment. *Cognition* **6**, 117–133.

Guindon, R. (1985) Anaphora Resolution: Short-Term Memory and Focusing. In *Proceedings of the 23rd Annual Meeting of the Association for Computational Linguistics*, pp. 218–227, University of Chicago.

Halliday, M. A. K. (1970) Language structure and language function. In J. Lyons, ed., *New Horizons in Linguistics.* Harmondsworth, Middlesex: Penguin.

Haviland, S. E. and H. H. Clark (1974) What's new? Acquiring new information as a process in comprehension. *Journal of Verbal Learning and Verbal Behavior* **13**, 512–521.

Hesiod (1959) *The works and days: Opera et dies.* Ann Arbor: University of Michigan Press. Translated by Richard Lattimore.

Howard, D. and S. Franklin (1990) Memory without rehearsal. In G. Vallar and T. Shallice, eds., *Neuropsychological impairments of short-term memory*, pp. 287–318. Cambridge: Cambridge University Press.

Hulme, C., S. Maughan and G. D. A. Brown (in presss) Memory for words and non-words: Evidence for a long term memory contribution to short-term memory tasks. *Journal of Memory and Language*.

Huttenlocher, J. (1968) Constructing spatial images: a strategy in reasoning. *Psychological Review* **75**, 550–560.

Huttenlocher, J., K. Eisenberg and S. Strauss (1968) Comprehension: Relation between Perceived Actor and Logical Subject. *Journal of Verbal Learning and Verbal Behavior* **7**, 527–530.

Huttenlocher, J. and S. Strauss (1968) Comprehension and a Statement's Relation to the Situation It Describes. *Journal of Verbal Learning and Verbal Behavior* **7**, 300–304.

Jakimik, J. and A. Glenberg (1990) Verbal Learning Meets Psycholinguistics: Modality Effects in the Comprehension of Anaphora. *Journal of Memory and Language* **29**, 582–590.

Johnson-Laird, P., D. Herrmann and R. Chaffin (1984) Only connections: A critique of semantic networks. *Psychological Bulletin* **2**, 292–315.

Johnson-Laird, P. N. (1983) *Mental Models.* Cambridge: Cambridge University Press.

Jones, G. V. (1976) A fragmentation hypothesis of memory: cued recall of pictures and of sequential position. *Journal of Experimental Psychology* **105**, 277–293.

Jones, G. V. (1979) Analysing memory by cuing: intrinsic and extrinsic knowledge. In N. S. Sutherland, ed., *Tutorial essays in psychology*, Vol. 2, pp. 119–145. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Jones, G. V. (1978) Tests of a structural theory of the memory trace. *British Journal of Psychology* **69**, 351–367.

Karmiloff-Smith, A. (1985) Language and Cognitive Processes from a Developmental Perspective. *Language and Cognitive Processes* **1**, 61–85.

Keenan, E. L. (1976) Toward a universal definition of "subject". In C. Li, ed., *Subject and Topic*, pp. 305–333. New York: Academic Press.

Kieras, D. E. (1978) Good and Bad Structure in Simple Paragraphs: Effects on Apparent Theme, Reading Time, and Recall. *Journal of Verbal Learning and Verbal Behavior* **17**, 13–28.

Kieras, D. E. (1981) Component processes in the comprehension of simple prose. *Journal of Verbal Learning and Verbal Behavior* **20**, 1–23.

Kintsch, W. (1988) The Role of Knowledge in Discourse Comprehension: A Construction-Integration Model. *Psychological Review* **95**, 163–182.

Kintsch, W. (1989) The representation of knowledge and the use of knowledge in discourse comprehension. In C. Graumann and R. Dietrich, eds., *Language in the Social Context.* Amsterdam: Elsevier.

Kolers, P. A. (1975) Specificity of operations in sentence recognition. *Cognitive Psychology* **1**, 283–306.

Kucera, H. and W. N. Francis (1967) *Computational Analysis of Present-Day American English.* Providence, Rhode Island: Brown University Press.

La Pointe, L. B. and R. W. Engle (1990) Simple and Complex Word Spans as Measures of Working Memory Capacity. *Journal of Experimental Psychology (Learning, Memory and Cognition)* **16**, 1118–1133.

Lesgold, A. M., S. F. Roth and M. E. Curtis (1979) Foregrounding effects in discourse comprehension. *Journal of Verbal Learning and Verbal Behavior* **18**, 291–308.

Levy, B. A. (1977) Reading: Speech and meaning processes. *Journal of Verbal Learning and Verbal Behavior* **16**, 623–638.

Levy, B. A. (1978) Speech analysis during sentence processing. Reading and listening. *Visible Language* **12**, 81–101.

Levy, J. and D. Bairaktaris (1991) A Model of the Interaction Between Long and Short Term Memory. In *International Joint Conference on Neural Networks*, pp. 1741–1746, Singapore.

Lewis, D. and C. J. Burke (1949) The use and misuse of the chi-square test. *Psychological Bulletin* **46**, 433–489.

Loftus, E. F. and P. Suppes (1972) Structural variables that determine the speed of retrieving words from long-term memory. *Journal of Verbal Learning and Verbal Behavior* **11**, 770–777.

Luria, A. R. (1968) *The Mind of a Mnemonist.* New York: Basic Books.

Maratsos, M. P. (1973) The effects of stress on the understanding of pronominal coreference in children. *Journal of Psycholinguistic Research* **2**, 1–8.

Masson, M. E. J. and L. S. Sala (1978) Interactive processes in sentences comprehension and recognition. *Cognitive Psychology* **10**, 244–270.

McClelland, J. L., D. E. Rumelhart and G. E. Hinton (1986) The Appeal of Parallel Distributed Processing. In *Parallel Distributed Processing: Exploration in the Microstructures of Cognition*, Vol. 1: *Foundations.* Cambridge, Mass.: MIT Press.

McKeithen, K. B., J. S. Reitman, H. H. Rueter and S. C. Hirtle (1981) Knowledge organization and skill differences in computer programmers. *Cognitive Psychology* **13**, 307–325.

Miller, G. A. (1962) Some psychological studies of grammar. *American Psychologist* **17**, 748–762.

Minsky, M. (1975) A Framework for Representing Knowledge. In P. Winston, ed., *The Psychology of Computer Vision*, pp. 211–277. New York: McGraw Hill.

Monsell, S. (1984) Components of working memory underlying verbal skills: A distributed capacities view — A tutorial review. In H. Bouma and D. G. Bouwhuis, eds., *Attention* *and Performance*, Vol. X, pp. 327–350. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Monsell, S. (1987) On the relation between lexical input and output pathways for speech. In *Language Perception and Production*, pp. 274–311. London: Academic Press Inc.

Morris, N. and D. Jones (1990) Memory updating in working memory: The role of the central executive. *British Journal of Psychology* **81**, 111–121.

Morris, P. E., M. M. Gruneberg, R. N. Sykes and A. Merrick (1981) Football knowledge and the acquisition of new results. *British Journal of Psychology* **72**, 479–483.

Morris, P. E., S. Jones and P. Hampson (1978) An imagery mnemonic for the learning of people's names. *British Journal of Psychology* **69**, 335–336.

Morris, P. E. and R. L. Reid (1970) The repeated use of mnemonic imagery. *Psychnomic Science* **20**, 337–338.

Morris, P. E. and R. Stevens (1974) Linking images and free recall. *Journal of Verbal Learning and Verbal Behavior* **13**, 310–315.

Moscovitch, M. and F. I. M. Craik (1976) Depth of processing, retrieval cues, and uniqueness of encoding as factors in recall. *Journal of Verbal Learning and Verbal Behavior* **15**, 447–458.

Murray, D. J. (1968) Articulation and acoustic confusability in short-term memory. *Journal of Experimental Psychology* **78**, 679–684.

Nebes, R. D. (1975) The nature of internal speech in a patient with aphemia. *Brain and Language* **2**, 489–497.

Neely, J. H. (1977) Semantic priming and retreival from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology: General* **106**, 226–254.

Nelson, A. W. R. (1988) A Simulation of Stereotypy in a Parallel Distributed Processing Framework. Master's thesis, Centre for Cognitive Science, University of Edinburgh.

Oakhill, J., A. Garnham and W. Vonk (1989) The On-Line Construction of Discourse Models. *Language and Cognitive Processes* **4**, 263–286.

Olson, G. M., R. L. Mack and S. A. Duffy (1981) Cognitive aspects of genre. *Poetics* **10**, 238–315.

Papagno, C., T. Valentine and A. Baddeley (1991) Phonological Short-Term Memory and Foreign-Language Vocabulary Learning. *Journal of Memory and Language* **30**, 331–347.

Perfetti, C. A. and A. M. Lesgold (1977) Discourse comprehension and sources of individual differences. In M. A. Just and P. A. Carpenter, eds., *Cognitive processes in comprehension*, pp. 141–183. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Posner, M. I. (1978) *Chronometric explorations of the mind.* Hillsdale, N. J.: Erlbaum.

Posner, M. I. and C. R. R. Snyder (1975) Attention and cognitive control. In R. L. Solso, ed., *Information Processing and Cognition: The Loyola Symposium.* Hillsdale, N.J.: Lawrence Erlbaum Associates.

Rayner, E. H. (1958) A study of evaluative problem solving. Part 1: Observations on adults. *Quarterly Journal of Psychology* **10**, 155–165.

Reisberg, D., I. Rappaport and M. O'Shaughnessy (1984) Limits of Working Memory: The Digit Digit-Span. *Journal of Experimental Psychology (Learning, Memory and Cognition)* **10**, 203–221.

Rips, L. J., E. J. Shoben and E. E. Smith (1973) Semantic Distance and the Verification of Semantic Relations. *Journal of Verbal Learning and Verbal Behaviour* **12**, 1–20.

Rondal, J. A., S. Brédart, N. Leyen, P. Neuville and F. Péree (1984) Coreference et strategie des fonctions paralleles dans le cas des pronoms anaphoriques ambihus. *Cahiers de Psychologie Cognitive* **4**, 151–170.

Ross, J. and K. A. Lawrence (1968) Some observations on memory artifice. *Psychonomic Science* **13**, 107–108.

Rumelhart, D. E. and D. A. Norman (1985) Representation of knowledge. In A. M. Aitkenhead and J. M. Slack, eds., *Issues in cognitive modelling*, pp. 15–62. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Sachs, T. S. (1974) Memory in reading and listening to discourse. *Memory and Cognition* **2**, 95–100.

Salamé, P. and A. D. Baddeley (1982) Disruption of short-term memory by unattented speech: Implications for the structure of working memory. *Journal of Verbal Learning and Verbal Behavior* **21**, 150–164.

Salzberg, P. M. (1976) On the generality of encoding specificity. *Journal of Experimental Psychology (Human Learning and Memory)* **2**, 586–596.

Sanford, A. J. and S. C. Garrod (1981) *Understanding Written Language.* Chichester: John Wiley and Sons.

Schacter, D. L. (1987) Implicit Memory: History and Current Status. *Journal of Experimental Psychology (Learning, Memory and Cognition)* **13**, 501–518.

Schank, R. C. (1972) Conceptual Dependency. *Cognitive Psychology* **3**, 552–631.

Schank, R. C. (1982) *Dynamic Memory: A Theory of Learning in People and Computers.* Cambridge: Cambridge University Press.

Sheldon, A. (1974) The role of parallel function in the acquisition of relative clauses in English. *Journal of Verbal Learning and Verbal Behavior* **13**, 272–281.

Shneiderman, B. (1976) Exploratory experiments in programmer behavior. *International Journal of Computer and Information Sciences* **5**, 123–143.

Slowiaczek, M. L. and C. C. Clifton (1980) Subvocalization and Reading for Meaning. *Journal of Verbal Learning and Verbal Behavior* **19**, 573–582.

Smith, E. E., E. J. Shoben and L. J. Rips (1974) Structure and process in semantic memory: a featural model for semantic decisions. *Psychological Review* **81**, 214–241.

Smith, K. H. and P. W. Foos (1975) Effect of presentation order on the construction of linear orders. *Memory and Cognition* **3**, 614–618.

Solan, L. (1983) *Pronominal Reference: Child language and the theory of grammar.* Dordrecht: D. Reidel.

Stenning, K. and J. Levy (1988) Knowledge-rich solutions to the 'binding problem': some human computational mechanisms. *Knowledge Based Systems* **1**.

Stenning, K., M. J. Patel and J. Levy (1987) The 'Binding Problem' in human memory: some effects of referential discontinuity on the construction of representations for individuals. Technical Report. Edinburgh University.

Stenning, K., M. Shepherd and J. Levy (1988) On the construction of representations for individuals from descriptions in text. *Language and Cognitive Processes* **2**, 129–164.

Stenning, K. (1986) On making models: a study of constructive memory. In T. Myers, K. Brown and B. McGonigle, eds., *Reasoning and Discourse Processes,* pp. 165–185. London: Academic Press.

Stenning, K. (1991) Binding attributes to individuals in terms of sequential information.

Stevenson, R. J. (1979) The effects of linguistic and cognitive factors on memory for sentences. Ph.D. thesis, University College, London.

Tehan, G. and M. S. Humphreys (1988) Articulatory loop explanations of memory span and pronunciation rate correspondences: A cautionary note. *Bulletin of the Psychonomic Society* **26**, 293–296.

Treisman, A. M. (1964) Verbal cues, language, and meaning in selective attention. *American Journal of Psychology* **77**, 206–219.

Tulving, E. (1972) Episodic and semantic memory. In *Organization of memory.* N.Y.: Academic Press.

Vallar, G. and S. F. Cappa (1987) Articulation and verbal short-term memory: Evidence from anarthia. *Cognitive Neuropsychology* **4**, 55–78.

Vallar, G. and T. Shallice (1990) *Neropsychological impairments of short-term memory.* Cambridge: Cambridge University Press.

Von Neumann, J. (1987) *Papers of John Von Neumann on computing and computing theory*, Vol. 12. Cambridge, Mass.: MIT Press.

Wanner, E. and M. Maratsos (1978) An ATN Approach to Comprehension. In M. Halle, J. Bresnan and G. A. Miller, eds., *Linguistic Theory and Psychological Reality*. Cambridge, Mass.: MIT Press.

Wetherick, N. E. (1975) The role of semantic information in short-term memory. *Journal of Verbal Learning and Verbal Behavior* **14**, 471–480.

Wickelgren, W. A. (1969) Context-sensitive Coding, Associative Memory, and Serial Order in (Speech) Behavior. *American Philosophical Quarterly* **76**, 1–15.

Wright, C. E. (1979) Duration differences between rare and common words and their implications for the interpretation of word frequency effects. *Memory and Cognition* **7**, 411–419.

Yates, F. A. (1966) *The art of memory*. London: Routledge and Kegan Paul.

Appendix A

Appendix to Chapter 2

A.1 Vocabulary sets used in Experiments I and II

A.1.1 Experiment I vocabulary

Dimension 1	Dimension 2	Dimension 3	Dimension 4
circle/square	black/white	old/new	thick/thin
triangle/oval	red/green	hard/soft	deep/shallow
rectangle/ellipse	yellow/blue	rough/smooth	hollow/solid
beam/block	silver/gold	wet/dry	large/small
cylinder/pyramid	bright/dim	hot/cold	long/short
disc/cube	shiny/dull	light/heavy	wide/narrow

Table A.1: Vocabulary set for Experiment I by dimension

A.1.2 Experiment II vocabulary

Dimension 1 Profession	Dimension 2 Nationality
secretary (74) / doctor (108)	Russian (34) / English(47)
judge (41) / maid (21)	Welsh (4) / Greek (20)
teacher (52) / clerk (22)	American (197) / French (58)
singer (12) / miner (7)	Canadian (6) / Polish (7)
nun (7) / vet (1)	Dutch (6) / Czech (-)
actress (7) / chef (7)	Portuguese(2) / Swiss (6)

Dimension 3 Personality	Dimension 4 Physical characteristics
generous (23) / modest (26)	gentle (24) / heavy (86)
bright (59) / sad (25)	young (187) / old (257)
alert (24) / dull (26)	healthy (21) / weak (23)
gloomy (3) / cheerful (10)	graceful (9) / awkward(8)
shy (11) / gruff (3)	slim (9) / stout (2)
eccentric (8) / sane (8)	hairy (5) / bald (4)

Table A.2: Vocabulary set for Experiment II by dimension. Numbers in parentheses are frequency of occurrence out of the 500 sample from Francis & Kucera's (1982) word count.

Appendix B

Appendix to Chapter 3

B.1 Materials used in pronoun comprehension experiments

Animate nouns				Inanimate nouns	
Noun	Sex	Noun	Sex	Noun	Noun
girl	f	schoolmistress	f	monitor	meter
secretary	f	nurse	f	palette	brush
sister	f	matron	f	spanner	wrench
mother	f	daughter	f	table	chair
doctor	m	dentist	m	lorry	car
actor	m	director	m	fridge	cooker
editor	m	journalist	m	bucket	rubbish
painter	m	sculptor	m	coffee	mug
councillor	m	mayor	m	fork	spoon
author	m	publisher	m	shirt	tie
singer	m	composer	m	sugar	milk
voter	m	candidate	m	film	camera
manager	m	assistant	m	salt	pepper
worker	m	boss	m	oil	petrol
president	m	chairperson	m	hammer	nail
priest	m	monk	m	spanner	bolt
juror	m	usher	m	shopping	bag
prisoner	m	warder	m	letter	tray
soldier	m	sailor	m	teapot	kettle
dustman	m	gardener	m	coal	scuttle
builder	m	architect	m	butter	jam
musician	m	conductor	m	beer	glass
customer	m	salesman	m	dishwasher	oven
burglar	m	constable	m	aerial	tv
judge	m	foreman	m	pen	ink
millionaire	m	pauper	m	match	box
pilot	m	designer	m	tobacco	pipe
lieutenant	m	general	m	spade	bucket
plumber	m	electrician	m	glass	bottle
thief	m	victim	m	ball	racket
trucker	m	hitchiker	m	paper	file
host	m	guest	m	battery	radio
gunman	m	hostage	m	arrow	bow
boy	m	schoolmaster	m	club	bag
curate	m	vicar	m	gun	silencer
captain	m	soldier	m	hat	coat
fireman	m	policeman	m	card	pack
bishop	m	cardinal	m	mug	teacup
chef	m	waiter	m	painting	sculpture
dustman	m	docker	m	hammer	toolbox
barman	m	drunk	m	mixture	bowl
guard	m	robber	m	staple	staplegun
undertaker	m	vicar	m	bullet	rifle
verger	m	deacon	m	slide	projector
mechanic	m	driver	m	ball	stick
farmer	m	vet	m	string	knife
jockey	m	trainer	m	ruler	rubber
manager	m	player	m	crayon	pencil

Table B.1: Animate and inanimate nouns used to make prototype sentences in Experiment III. The sex of the animate nouns is included (m = male, f = female). Animate and inanimate nouns on the same line made up a pair of nouns which always appeared together.

Table B.2: Verb phrase frames used in Experiment III.

First clause		Second Clause		Noun	
Verb	PP	Verb	PP	order	Question (True/False)
put	near to	shoved	opposite to	ab	facing
					beside
moved	to the right of	placed	behind	ba	in front of
					to the rear of
shoved	nearby	placed	to the left of	ab	to the left of
					to the right of
shoved	to the left of	placed	further away from	ba	far from
					near to
moved	beside	positioned	near to	ab	close to
					far away from
moved	behind	shoved	to the left of	ba	to the right of
					to the left of
positioned	to the right of	put	close to	ab	near to
					far from
moved	nearby	pushed	in front of	ba	behind
					in front of
					far from
positioned	opposite	placed	nearby	ba	close to
					far from
positioned	close to	placed	behind	ab	to the rear of
					in front of
shoved	to the right of	put	beside	ba	beside
					opposite to
put	opposite to	placed	near to	ab	close to
					far away from
put	behind	pushed	close to	ba	near to
					far from
put	close to	placed	nearby	ab	near to
					far from
positioned	further away from	shoved	to the left of	ba	to the right of
					to the left of
pushed	to the left of	positioned	further away from	ab	far away from
					near to
put	opposite	shoved	in front of	ba	behind
					in front of
shoved	next to	placed	to the right of	ab	to the right of
					to the left of
pushed	to the left of	positioned	close to	ba	near to
					far away from
placed	close to	shoved	opposite to	ab	facing

					beside
pushed	to the left of	moved	behind	ba	in front of to the rear of
put	behind	positioned	to the left of	ab	to the left of to the right of
pushed	near to	positioned	opposite	ba	opposite beside
put	close to	placed	beside	ab	near to far from
moved	opposite to	put	in front of	ba	beahind in front of
put	opposite to	pushed	nearby	ab	near to far from
pushed	near to	moved	to the left of	ba	to the right of to the left of
put	next to	pushed	to the left of	ab	to the left of to the right of
moved	opposite	placed	nearby	ba	close to far from
shoved	further away from	placed	in front of	ab	in front of behind
positioned	further away from	placed	opposite to	ba	opposite besdie
positioned	further away from	placed	to the left of	ab	to the left of to the right of
put	further away from	shoved	opposite	ba	opposite beside
pushed	opposite to	moved	close to	ab	near far from
put	to the right of	shoved	further away from	ba	far from near to
placed	nearby	moved	beside	ab	near to far from
moved	nearby	pushed	in front of	ba	behind in front of
put	to the right of	placed	near to	ab	close to far from
placed	beside	shoved	to the left of	ba	to the right of to the left of
put	near to	pushed	behind	ab	to the rear of in front of
placed	near to	put	to the right of	ba	to the left of to the right of
moved	further away from	pushed	opposite to	ab	facing

					beside
put	in front of	placed	close to	ba	near to
					far from
moved	opposite to	shoved	nearby	ab	near to
					far from
moved	close to	shoved	to the right of	ba	to the left of
					to the right of
pushed	nearby	shoved	to the right of	ab	to the right of
					to the left of
shoved	nearby	pushed	next to	ba	next to
					far away from

		Animate nouns				Inanimate nouns		
Noun	Sex	Noun	Sex	Noun	Sex	Noun	Noun	Noun
girl	f	prefect	f	schoolmistress	f	monitor	meter	dial
secretary	f	nurse	f	doctor	m	palette	brush	knife
sister	f	matron	f	patient	f	spanner	wrench	bolt
mother	f	daughter	f	father	m	table	chair	lamp
doctor	m	dentist	m	hygenist	f	lorry	car	bus
actor	m	stage-hand	m	director	m	fridge	cooker	twintub
journalist	m	cartoonist	m	editor	m	bucket	rubbish	cooker
painter	m	sculptor	m	model	f	coffee	mug	teaspoon
councillor	m	mayor	m	vicar	m	fork	spoon	knife
publisher	m	editor	m	author	m	shirt	tie	jacket
composer	m	conductor	m	singer	m	sugar	milk	biscuit
candidate	m	canvasser	m	voter	m	film	camera	lens
salesman	m	man	m	manager	m	salt	pepper	ketchup
worker	m	shop steward	m	boss	m	oil	petrol	anti-freeze
president	m	chairman	m	secretary	m	hammer	nail	screwdriver
priest	m	monk	m	cardinal	m	spanner	bolt	wrench
juror	m	usher	m	judge	m	shopping	bag	trolley
prisoner	m	warder	m	governor	m	letter	tray	knife
soldier	m	sailor	m	provost	m	teapot	kettle	tea caddy
dustman	m	gardener	m	driver	m	coal	scuttle	poker
builder	m	foreman	m	architect	m	butter	jam	bread
musician	m	conductor	m	soloist	m	beer	glass	wine
customer	m	salesman	m	manger	m	dishwasher	oven	fridge
constable	m	sergeant	m	burglar	m	aerial	tv	video
judge	m	foreman	m	witness	m	pen	ink	paper
pauper	m	banker	m	millionaire	m	match	box	candle
pilot	m	designer	m	mechanic	m	tobacco	pipe	pipe cleaner
lieutenant	m	general	m	private	m	spade	bucket	trowel
plumber	m	electrician	m	plasterer	m	glass	bottle	beermat
thief	m	victim	m	constable	m	ball	racket	net
trucker	m	hitchiker	m	mechanic	m	paper	file	holepuncher
guest	m	caterer	m	host	m	battery	radio	earpiece
gunman	m	hostage	m	pilot	m	arrow	bow	target
prefect	m	boy	m	schoolmaster	m	club	bag	ball
curate	m	vicar	m	verger	m	gun	silencer	bullet
soldier	m	sapper	m	captain	m	hat	coat	scarf
fireman	m	policeman	m	bystander	m	card	pack	dice
bishop	m	priest	m	cardinal	m	teacup	saucer	mug
chef	m	waiter	m	cook	m	painting	sculpture	easel
dustman	m	docker	m	engineer	m	hammer	saw	toolbox
barman	m	drunk	m	doorman	m	flour	egg	bowl
guard	m	thief	m	policeman	m	staple	staplegun	paper
undertaker	m	vicar	m	mourner	m	bullet	rifle	sling
verger	m	deacon	m	prebendry	m	slide	projector	cartridge
mechanic	m	driver	m	passenger	m	ball	stick	hoop
farmer	m	vet	m	farmhand	m	string	knife	stapler
jockey	m	trainer	m	owner	m	ruler	rubber	sharpener
manager	m	player	m	trainer	m	crayon	pencil	paper

Table B.3: Animate and inanimate noun triples used to make prototype sentences in Experiment VI. The sex of the animate nouns is included after each noun (m = male, f = female). Each line of the table made up one prototype sentence and the first two nouns of the animate and inanimate triples appeared in the first clause and the remaining noun appeared in the second clause.

	Animate			Inanimate	
Noun1	Noun2	Noun3	Noun1	Noun2	Noun3
schoolmistress	boy	prefect	monitor	meter	dial
nurse	porter	doctor	palette	brush	knife
matron	anaesthetist	surgeon	spanner	wrench	bolt
mother	son	father	table	chair	lamp
hygienist	dentist	patient	car	bus	lorry
actress	stage-hand	director	fridge	cooker	twintub
typist	journalist	editor	bin	rubbish	cooker
model	painter	dealer	mug	coffee	teaspoon
mayoress	councillor	vicar	fork	spoon	knife
authoress	publisher	editor	shirt	tie	jacket
soprano	composer	conductor	biscuit	sugar	milk
housewife	candidate	reporter	film	camera	lens
woman	salesman	manager	brick	shovel	barrow
tea-lady	shop steward	boss	spoon	plate	milk
secretary	president	chairman	tape	record	tuner
nun	priest	monk	paint	brush	wallpaper
stripper	usher	judge	pipe	gearbox	clamp
manageress	detective	salesman	hammer	nail	screwdriver
millionairess	pauper	banker	screw	chisel	drill
cleaning woman	janitor	executive	bag	trolley	shopping
hostess	waiter	caterer	letter	tray	knife
au pair	gardener	chauffeur	teapot	kettle	tea caddy
sempstress	tailor	gentleman	poker	coal	scuttle
maid	butler	valet	dishwasher	oven	fridge
milkmaid	farmhand	farmer	mvideo	aerial	tv
chambermaid	bellboy	manager	pen	ink	paper
landlady	lodger	husband	candle	match	box
duchess	equerry	page-boy	pipe	tobacco	pipe cleaner
governess	boy	schoolmaster	bucket	trowel	spade
poetess	publisher	critic	bottle	beermat	glass
ballerina	director	stage-hand	ball	racquet	net
heiress	suitor	gigolo	file	holepuncher	paper
millionairess	banker	beggar	radio	earpiece	battery
barmaid	customer	manager	arrow	bow	target
stripper	man	policeman	club	bag	ball
headmistress	parent	boy	bullet	gun	silencer
nannie	master	gardener	hat	coat	scarf
stewardess	passenger	pilot	card	pack	dice
usherette	patron	projectionist	teacup	saucer	mug
au pair	agent	housekeeper	painting	sculpture	easel
contralto	chorister	organist	hammer	saw	toolbox
marchioness	courtier	judge	staple	staplegun	paper
air hostess	steward	passenger	bullet	rifle	sling
receptionist	cabbie	doorman	slide	projector	cartridge
waitress	waiter	diner	ball	stick	hoop
woman	lawyer	judge	paperclip	string	stapler
waitress	barman	owner	ruler	rubber	sharpener
coiffeuse	hippy	mother	crayon	pencil	paper

Table B.4: Animate and Inanimate nouns used in Experiment VII. For each row, Nouns' 1 and 2 appeared in the first clause and Noun3 appeared in the second clause. Of the animate nouns, all those in column Noun1 were treated as feminine, all those in column Noun2 were treated as masculine and all those in column Noun3 were treated as masculine except for *housekeeper* and *mother*.

NP_1	NP_2
schoolmistress	girl
spectator	player
comedian	compere
constable	motorist
priest	widower
captain	soldier
guard	suspect
matron	nurse
golfer	professional
warder	convict
bowler	umpire
photographer	policeman
editor	reporter
headmaster	parent
teacher	'new-boy'
captain	sailor
assistant	customer
judge	lawyer
dentist	patient
lecturer	student
vet	farmer
witness	prisoner
client	salesman
dealer	buyer
pianist	singer
bishop	parishioner
gamekeeper	poacher
director	actor
doctor	patient
father	boy
sergent	private
passenger	driver

Table B.5: Nouns used to make up protype sentences in Experiment IV.

	Agentive			Non-agentive	
VP_1	VP_2	Adv P	VP_1	VP_2	Adv P
punished	apologized to	much later	heard	surprised	by being there
blamed	insulted	after the match	saw	cheered	up by being there
bored	disliked	from then on	respected	amazed	by his act
cautioned	drove past	in a hurry	noticed	surprised	by his presence
comforted	talked to	for a while	understood	talked to	for a while
encouraged	trusted	with his life	heard	told	about the war
frisked	recognized	immediately	noticed	threatened	with violence
harassed	avoided	on the ward	liked	visited	at weekends
challenged	beat	on the course	envied	beat	on the course
released	insulted	behind his back	distrusted	frightened	with his threats
obeyed	commended	after the game	respected	agreed with	after the game
defied	apologised to	after the demo	saw	chased after	with great caution
criticized	thanked	anyway	noticed	offered	a story
reported to	thanked	for the meeting	respected	thanked	for his concern
welcomed	befriended	quickly	resented	disliked	as well
reprimanded	liked	after that	distrusted	insulted	on the bridge
insulted	made	very angry	noticed	ignored	purposefully
insulted	avoided	after the trial	despised	insulted	with good reason
examined	explained to	about the pain	heard	apologised to	after the treatment
spoke to	avoided	for the next week	saw	avoided	for the next week
telephoned	delayed	with old gossip	liked	encouraged	with new results
accused	hated	for lying	saw	hated	with good reason
telephoned	argued with	over the price	distrusted	persuaded	to start haggling
congratulated	settled with	later on	believed	trusted	as well
hurried	helped	in the last verse	appreciated	rewarded	with a contract
scolded	forgave	afterwards	heard	chatted to	after mass
caught	assaulted	needlessly	heard	shot	needlessly
auditioned	thanked	quickly	saw	approached	much later
helped	befriended	quite quickly	respected	talked to	for a while
scolded	apologized to	soon afterwards	believed	trusted	after that
bullied	insulted	off parade	distrusted	surprised	in action
criticised	loathed	throughout the journey	liked	talked to	throughout the journey

Table B.6: Agentive and non-agentive verbs and associated adverbial phrases in Experiment IV.

Table B.7: Illustrative examples of materials used in Experiment V and in Experiment VIII. Each entry contains three pairs of context sentences and a target sentence. The three contexts correspond to Conjoined, Subject-Predicate and Separate levels of Text. In the actual experiment each context was followed by the target and in the Single level of Text the target was presented in isolation. The examples all contain targets with Subject Pronouns and in the Same Order.

John and Sammy were playing in the garden.
Ellen watched their game with interest.

John was playing with Sammy in the garden.
Ellen watched their game with interest.

John and Ellen were playing in the garden.
Sammy watched their game with interest.

John pushed Sammy and he kicked Ellen.

Brenda and Harriet were starring in the local musical.
Bill was in it too and none of them were very sure of their dance steps.

Brenda was co-starring with Harriet in the local musical.
Bill was in it too and none of them were very sure of their dance steps.

Brenda and Bill were starring in the local musical.
Harriet was in it too and none of them were very sure of their dance steps.

Brenda copied Harriet and she watched Bill.

Mary and Priscilla were about to go into town when they realised the car had a puncture.
Graham was their next door neighbour and he knew nothing about cars.

Mary was about to go into town with Priscilla when they realised the car had a puncture.
Graham was their next door neighbour and he knew nothing about cars.

Mary and Graham were about to go into town when they realized the car had a puncture.
Priscilla was their next door neighbour and she knew a bit about cars.

Mary helped Priscilla change the wheel and she talked to Graham.

Shirley and Vanessa were organizing the Christmas pantomime.
Martin had agreed to help with the production.

Shirley was co-producing the Christmas pantomime with Vanessa.
Martin had agreed to help with the production.

Shirley and Martin were organizing the Christmas pantomime.
Vanessa had agreed to help with the production.

Shirley wrote to Vanessa about a meeting and she phoned Martin.

Michael and Philip were at a cricket match.
Their friend Helen came along to watch as well.

Michael was going to meet Philip at the cricket match.
Their friend Helen came along to watch as well.

Michael and Helen were at a cricket match.
Their friend Philip came along to watch as well.

Michael took Philip to the pavilion and he waved to Helen.

Richard and Jim were playing cops and robbers on the old playing fields.
Their classmate Caroline passed by on he way to the shops.

Richard was playing cops and robbers with Jim on the old playing fields.
Their classmate Caroline passed by on he way to the shops.

Richard and Caroline were playing cops and robbers on the old playing fields.
Their classmate Jim was keen to play a game.

Richard chased Jim round the corner and he ignored Caroline.

Malcolm and Arthur were always competing at work.
Gillian thought it was getting very boring.

Malcolm was always competing with Arthur at work.
Gillian thought it was getting very boring.

Malcolm and Gillian were always competing at work.
Arthur thought it was getting very boring.

Malcolm criticised Arthur and he scowled at Gillian.

Robert and Peter used to be best friends.
Freda was their friend as well and she always made things worse when they quarrelled.

Robert used to be Peter's best friend.
Freda was their friend as well and she always made things worse when they quarrelled.

Robert and Freda used to be best friends.
Peter was their friend as well and he always made things worse when they quarrelled.

Robert bullied Peter and he attacked Freda.

Vincent and Ken lived on the same street and arranged to go into work togethe.
Julie was cross because they never offered he a lift.

Vincent lived on the same street as Ken and they arranged to go into work togethe.
Julie was cross because they never offered he a lift.

Vincent and Julie lived on the same street and arranged to go into work togethe.
Ken wanted leave at a different time and was very insistent.

Vincent drove Ken to town and he glared at Julie.

Rachel and Suzanne were jealous of each other.
Bob worked with them and had noticed that they didn't get on.

Rachel was jealous of Suzanne and it was clear that the feeling was mutual.
Bob worked with them and had noticed that they didn't get on.

Rachel and Bob were working on the same project.
Suzanne also worked with them as an advisor.

Rachel criticised Suzanne and she sympathized with Bob.

Elizabeth and Sue had just had an argument and neitshe wanted to be the first to make up.
Steve was upset because he liked them both.

Elizabeth had just had an argument with Sue and neitshe wanted to be the first to make up.
Steve was upset because he liked them both.

Elizabeth had just had a great idea which Steve liked and wanted to hear more about.
Sue was interested but had some questions.

Elizabeth talked to Sue and she encouraged Steve.

Arnold and Gerald were finishing their drinks.
Their sister Tina had just come into the bar.

Arnold was having a long chat with Gerald and they'd both just finished their drinks.
Their sister Tina had just come into the bar.

Arnold and Tina were finishing their drinks.
Their brother Gerald had just come into the bar.

Arnold asked Gerald to get the next round and he laughed at Tina.

Geoffrey and Mark were lazing in the sun by the swimming pool.
Sara had some work to do and was sitting in the shade.

Geoffrey joined Mark lazing in the sun by the swimming pool.
Sara had some work to do and was sitting in the shade.

Geoffrey and Sara were lazing in the sun by the swimming pool.
Mark was feeling a bit restless.

Geoffrey followed Mark into the pool and he watched Sara.

Nigel and Sam were playing in the attic.
Maeve could hear the noise and went to find out what was going on.

Nigel went up to the attic to play with Sam.
Maeve could hear the noise and went to find out what was going on.

Nigel and Maeve were playing in the attic.
Sam could hear the noise and went to find out what was going on.

Nigel locked Sam in the old wardrobe and he heard Maeve.

Liz and Melanie were always fighting in the playground.
Frank often joined in when he got a chance.

Liz was always fighting with Melanie in the playground.
Frank often joined in when he got a chance.

Liz and Frank were always fighting in the playground.
Melanie often joined in when she got a chance.

Liz tried to catch Melanie and she chased Frank.

Dominic and Don decided to go boating on the river but they hadn't fixed on a meeting place.
Ruth wanted to go too and went to try to find them.

Dominic had arranged with Don to go boating on the river but they hadn't said wDe.
Ruth wanted to go too and went to try to find them.

Dominic had decided to go boating on the river with Ruth who came separately.
Don was coming too and arrived late.

Dominic was searching for Don and he saw Ruth.

Gloria and Emma went to their new friends birthday party.
Mike went too and decided they should have a breath holding contest.

Gloria invited Emma to she eighth birthday party.
Mike went too and decided they should have a breath holding contest.

Gloria and Mike went to their new friends birthday party.

Emma went too and decided they should have a breath holding contest.

Gloria tickled Emma and she laughed at Mike.

Stuart and James were both representing their school in the county athletics competition.
Jessica was also competing.

Stuart was in James's team in the county athletics competition.
Jessica was also competing.

Stuart and Jessica were both representing their school in the county athletics competition.
James was also competing.

Stuart raced James along the track and he watched Jessica.

Alison and Eleanor were sisters but they didn't get on very well.
Their elder brother Duncan was always pleased when they didn't fight.

Alison didn't get on very well with she sister Eleanor.
Their elder brother Duncan was always pleased when they didn't fight.

Alison and Duncan got on surprisingly well for brother and sister.
Their elder sister Eleanor was always a bit more awkward.

Alison asked Eleanor to dinner and she encouraged Duncan.

Patricia and Martha hadn't seen each other for years.
They arranged a reunion and invited their old friend Nicholas along.

Patricia hadn't seen she friend Martha for years.
They arranged a reunion and invited their old friend Nicholas along.

Patricia and Nicholas hadn't seen each otshe for years.
They arranged a reunion and invited their old friend Martha.

Patricia gave Martha a present and she smiled at Nicholas.

Sophie and Lynn decided that their club house needed decorating.
Simon thought it would cost too much.

Sophie persuaded Lynn that their club house needed decorating.
Simon thought it would cost too much.

Sophie tried to persuade Simon that their club house needed decorating.
Lynn thought it was getting too shabby as well.

Sophie helped Lynn paint the ceiling in the bar and she watched Simon.

Miranda and Pandora were great rivals and often fell out.

Daniel was fed-up with their squabbles but he sometimes took sides.

Miranda was a great rival of Pandora's and they often fell out.
Daniel was fed-up with their squabbles but he sometimes took sides.

Miranda and Daniel were great rivals and often fell out.
Pandora was fed-up with their squabbles but she sometimes took sides.

Miranda ignored Pandora and she spurned Daniel.

Max and Sean went for a game of snooker in their local sports centre.
Kate went along later so how they were doing.

Max met Sean at the local sports centre for a game of snooker.
Kate went along later just to watch.

Max and Kate went for a game of snooker in their local sports centre.
Sean went along later for a quick game.

Max played with Sean and he cheered Kate on.

Alice and Madeline went to the same chip shop every night.
Rex worked in the chip shop most nights and was glad to see them.

Alice met Madeline at the chip shop most nights.
Rex worked in the chip shop every night and was glad to see them.

Alice usually went to the chip ship where Rex worked.
Madeline went tDe every evening at the same time.

Alice let Madeline go first and she smiled at Rex.

Tom and Jeremy were organizing their staff party.
Julia pretended to be helping but she didn't really do much.

Tom was helping Jeremy organize their staff party.
Julia pretended to be helping but she didn't really do much.

Tom and Julia were tired of organizing the entire staff party.
Jeremy was always ready to help out.

Tom helped Jeremy to prepare the food and he chatted to Julia aimlessly.

Isabelle and Chloe went to the bank to arrange a business loan.
Reginald, the manager, asked them how much they wanted to borrow.

Isabelle went to the bank with she partner Chloe to arrange a business loan.
Reginald, the manager, asked them how much they wanted to borrow.

Isabelle went to see Reginald at the bank to arrange a business loan.
she partner Chloe had to go as well.

Isabelle kept quiet while Chloe talked and she frowned at Reginald disapprovingly.

Sarah and Cathy hadn't seen each other since leaving home years ago.
Their brother Charles also hadn't been in touch with eitshe of them for some time.

Sarah had just moved away from home and hadn't seen she sister Cathy for ages.
Their brother Charles also hadn't been in touch with eitshe of them for some time.

Sarah and Charles hadn't seen each otshe since leaving home years ago.
Their sister Cathy also hadn't been in touch with eitshe of them for some time.

Sarah visited Cathy at home and she rang Charles at work.

Bernard and Fraser were neighbours but they didn't really get on very well.
Louise lived round the corner and always wanted to know what was happening.

Bernard lived next door to Fraser but they didn't really get on very well.
Louise lived round the corner and always wanted to know what was happening.

Bernard and Louise were neighbours and they always wanted to know what was going on.
Fraser lived round the corner and didn't get on with people very well.

Bernard argued with Fraser and he asked Louise about it.

Anthony and Derek were going to present an important document at the staff meeting.
Carol had agreed to give a hand as well.

Anthony was going to help Derek present an important document at the staff meeting.
Carol had agreed to give a hand as well.

Anthony had been chosen by Carol to make a presentation at the staff meeting.
Derek had agreed to give a hand as well.

Anthony encouraged Derek and he asked Carol how it was going.

Jean and Penelope were working late trying to finish a job.
Roger was also working late but at nine o'clock had had enough.

Jean was working late with Penelope trying to finish a job.
Roger was also working late but at nine o'clock had had enough.

Jean was working late with Roger who was getting tired.
Penelope was also working on the same project.

Jean asked Penelope to carry on and she told Roger to have a break.

Claire and Beverley usually went into town together on Saturdays.
They sometimes stopped to see Bernie, their old headmaster.

Claire usually went into town on Saturdays with Beverley.
They sometimes stopped to see Bernie, their old headmaster.

Claire and Bernie usually went into town together on Saturdays.
They sometimes stopped to see Beverley, their old headmistress.

Claire envied Beverley and she warned Bernie about it.

Catriona and Maureen always exchanged presents at Christmas.
Their brother Joe often bought presents too.

Catriona always exchanged presents with Maureen at Christmas.
Their brother Joe often bought presents too.

Catriona and Joe always exchanged presents at Christmas.
Their sister Maureen usually exchanged presents too.

Catriona gave Maureen a new watch and she bought Joe a record.

Adam and Barry worked very competitively for the same newspaper.
Linda had just started working with them and was still new to the job.

Adam worked competitively with Barry on the same newspaper.
Linda had just started working with them and was still new to the job.

Adam and Linda worked very competitively for the same newspaper.
Barry had just started working with them but was rather opinionated.

Adam argued with Barry about the new layout and he questioned Linda about it.

Ewan and William had been close friends at school but were now parting to go away to college.
Fiona had known them both for years and organised a leaving party for them.

Ewan had been a close friend of William's at school and they were now parting to go away to college.
Fiona had known them both for years and organised a leaving party for them.

Ewan and Fiona had been close friends at school but were now parting to go away to college.
William had known them both for years and organised a leaving party for them.

Ewan gave William a lift to the party and he asked Fiona to open the wine.

David and Christoper were on holiday in the Lake District.
Daphne the landlady took an instant dislike to them.

David was on holiday in the Lake District with Christoper.
Daphne the landlady took an instant dislike to them.

When David was on holiday he stayed in a terrible hotel run by a woman called Daphne.
Christoper was staying in the same hotel.

David showed Christoper a leaflet about anotDaphne hotel and he glared at Daphne angrily.

Cheryl and Monica were members of the local peace group.
Steven had just joined and wasn't very involved yet.

Cheryl regularly went to the local peace group with Monica.
Steven had just joined and wasn't very involved yet.

Cheryl and Steven had been members of the local peace group for some time.
Monica had just joined and was very enthusiastic.

Cheryl spoke to Monica about the meeting and she asked Steven some questions.

Betty and Sally had been pestering the firm for months to buy a new photocopier.
Donald was hoping that he would be able to use it too.

Betty had been pestering Sally for months to buy the firm a new photocopier.
Donald was hoping that he would be able to use it too.

Betty and Donald had been pestering the firm for months to buy a new photocopier.
Sally was keen to use it too.

Betty demonstrated the new machine to Sally and she asked Donald about it.

Keith and Timothy were friends and often shared things with each othe.
They liked Hazel but often forgot about he.

Keith was friendly with Timothy and they often shared things with each othe.
They liked Hazel but often forgot about he.

Keith and Hazel were friends and often shared things with each othe.
They liked Timothy who was keen join their gang.

Keith shared his chocolate bar with Timothy and he asked Hazel for a piece.

Ronnie and Iain had big plans for the company's expansion.
Jenny wanted to know what was happening so she went to their meeting.

Ronnie had big plans for the company's expansion which Iain disagreed with.
Jenny wanted to know what was happening too so she went to their meeting.

Ronnie had big plans for the company's expansion which Jenny was interested in.

Iain was worried about the possible costs and he could easily halt the project.

Ronnie discussed things with Iain and he listened to Jenny closely.

Harry and Oscar were both applying for the same job.
Rosie had decided to hire them both.

Harry was applying for the same job as Oscar.
Rosie had decided to hire them both.

Harry was applying for a job with Rosie's company.
Oscar decided to apply for the same job.

Harry praised Oscar and he was pleased with Rosie.

Ben and Patrick loved to go swimming when the sea was rough.
Anna had noticed the strong undertow.

Ben loved to go swimming with Patrick when the sea was rough.
Anna had noticed the strong undertow.

Ben and Anna loved to go swimming when the sea was rough.
Patrick was watching from the clifftop.

Ben waved to Patrick and he shouted a warning to Anna.

Lorna and Gail went to the basketball match.
Their friend Paul was playing for the home side.

Lorna went with Gail to the basketball match.
Their friend Paul was playing for the home side.

Lorna went to watch Paul play for the local basketball team.
Their friend Gail was also at the match.

Lorna warned Gail about the side's poor performance and she surprised Paul with his skill.

Edward and Jack did a little gardening at weekends for extra pocket money.
Lucy decided to employ them for a few hours.

Edward helped Jack do a little gardening at weekends for extra pocket money.
Lucy decided to employ them for a few hours.

Edward and Lucy did a little gardening at weekends for extra pocket money.
Jack helped out on a job as a favour to them.

Edward started a bonfire while Jack weeded and he watched Lucy from the shed.

Elsie and Mavis liked to go sailing at weekends.

Hector often let them crew his boat.

Elsie liked to go sailing at weekends and often took Mavis along.
Hector often let them crew his boat.

Hector took Elsie sailing at weekends.
Mavis liked to come along too.

Elsie helped Mavis launch the boat and she told Hector what to do.

Archie and Larry only went hill-walking for the exercise.
Jill went along for the scenery as well.

Archie liked racing Larry when they went hill-walking.
Jill went along for the scenery.

Archie and Jill only went hill-walking for the exercise.
Larry went along for the scenery as well.

Archie walked beside Larry and he asked Jill to slow down.

Agnes and Daisy were thinking about going to see the new film at the cinema.
Len had seen it a few nights ago.

Agnes was discussing with Daisy whetshe to see to see the new film at the cinema.
Len had seen it a few nights ago.

Agnes was thinking of going to see the cinema's new film which Len had just seen.
Daisy also wanted to see it.

Agnes told Daisy that she'd seen a review and she warned Len it was rubbish.

Joan and Lillith had just managed to catch the evening train.
Mathew had already got a seat at a free table.

Joan was travelling home with Lillith and they just managed to catch the evening train.
Mathew had already got a seat at a free table.

Joan and Mathew had just managed to catch the evening train.
Their friend Lillith spotted some seats with a table.

Joan sat beside Lillith at the table and she started talking to Mathew.

Bella and Maxine had just arrived at the health farm.
Dougal had already started on his exercise regime and was looking very ill.

Bella had just arrived with Maxine at the health farm.
Dougal had already started on his exercise regime and was looking very ill.

Bella had just arrived at the health farm and was watching Dougal doing some exercise. Maxine was also looking about while she checked in.

Bella looked at Maxine apprehensively and she warned Dougal about the diet.

B.2 Analyses of question answers in pronoun comprehension experiments

Antecedent	Pronoun	Accuracy Correct	Incorrect	TOTAL
Subject	Subject	105	39	144
	Non-subject	84	60	144
	TOTAL	189	99	288
Non-subject	Subject	74	70	144
	Non-subject	75	69	144
	TOTAL	149	139	288

Table B.8: Frequency of correct and incorrect responses to question by Antecedent and Pronoun in Experiment III.

Text	Animacy	Antecedent	Pronoun	Accuracy		
				Correct	Incorrect	TOTAL
Dynamic	Animate	Subject	Subject	34	2	36
			Non-subject	26	10	36
			TOTAL	60	12	72
		Non-subject	Subject	22	14	36
			Non-subject	3	33	36
			TOTAL	25	47	72
	Inanimat	Subject	Subject	5	31	36
			Non-subject	21	15	36
			TOTAL	26	46	72
		Non-subject	Subject	23	13	36
			Non-subject	4	32	36
			TOTAL	27	45	72
Static	Animate	Subject	Subject	31	5	36
			Non-subject	18	18	36
			TOTAL	49	23	72
		Non-subject	Subject	18	18	36
			Non-subject	33	3	36
			TOTAL	51	21	72
	Inanimat	Subject	Subject	35	1	36
			Non-subject	19	17	36
			TOTAL	54	18	72
		Non-subject	Subject	11	25	36
			Non-subject	35	1	36
			TOTAL	46	26	72

Table B.9: Number of correct and incorrect responses by Animacy, Text, Antecedent and Pronoun in Experiment III.

Antecedent	Pronoun	Accuracy		
		Correct	Incorrect	TOTAL
Subject	Subject	107	37	144
	Non-subject	102	42	144
	TOTAL	209	79	288
Non-subject	Subject	98	46	144
	Non-subject	116	28	144
	TOTAL	214	74	288

Table B.10: Frequency of correct and incorrect responses to question by Antecedent and Pronoun in Experiment VII.

Text	Animacy	Antecedent	Pronoun	Accuracy Correct	Incorrect	TOTAL
Dynamic	Animate	Subject	Subject	30	6	36
			Non-subject	23	13	36
			TOTAL	53	19	72
		Non-subject	Subject	22	14	36
			Non-subject	29	7	36
			TOTAL	51	21	72
	Inanimate	Subject	Subject	21	15	36
			Non-subject	28	8	36
			TOTAL	49	23	72
		Non-subject	Subject	24	12	36
			Non-subject	29	7	36
			TOTAL	53	19	72
Static	Animate	Subject	Subject	27	9	36
			Non-subject	26	10	36
			TOTAL	53	19	72
		Non-subject	Subject	24	12	36
			Non-subject	32	4	36
			TOTAL	56	16	72
	Inanimate	Subject	Subject	29	7	36
			Non-subject	25	11	36
			TOTAL	54	18	72
		Non-subject	Subject	28	8	36
			Non-subject	26	10	36
			TOTAL	54	18	72

Table B.11: Number of correct and incorrect responses by Animacy, Text, Antecedent and Pronoun in Experiment VII.

	Accuracy		
Agency	Incorrect	Correct	TOTAL
Agent	59	133	192
Nonagent	92	100	192
TOTAL	151	233	384

Table B.12: Number of correct and incorrect question answers by Agency in Experiment IV.

Antecedent	Pronoun	Accuracy		
		Incorrect	Correct	TOTAL
Subject	Subject	43	53	96
	Non-subject	37	59	96
	TOTAL	80	112	192
Non-subject	Subject	35	61	96
	Non-subject	36	60	96
	TOTAL	71	121	192

Table B.13: Number of correct and incorrect questions answered by Antecedent and Pronoun in Experiment IV.

Appendix C

Appendix to Chapter 4

C.1 The vocabuary set used in Experiment IX

Dimension 1 Profession	Dimension 2 Nationality	Dimension 3 Personality	Dimension 4 Physical characteristics
dentist/baker	Chinese/Polish	bright/dumb	cheerful/sad
dentist/chef	Chinese/Swiss	clever/dim	daring/meek
doctor/priest	Dutch/Czech	clever/stupid	daring/timid
doctor/vicar	Dutch/Russian	clever/thick	friendly/hostile
judge/bishop	French/Greek	fat/skinny	happy/gloomy
judge/monk	French/Spanish	fat/thin	happy/sad
nurse/priest	German/Greek	greedy/clumsy	nice/nasty
nurse/vicar	German/Spanish	hungry/thirsty	normal/mad
teacher/bishop	Swedish/Czech	rich/poor	sane/insane
teacher/monk	Swedish/Russian	smart/scruffy	sane/mad
vet/baker	Welsh/Polish	wealthy/poor	strong/weak
vet/chef	Welsh/Swiss	young/old	tall/short

Experimental vocabulary

Table C.1: Vocabulary by dimension from Experiment IX

Appendix D

Appendix to Chapter 5

D.1 Vocabulary used in Experiment X

Dim.	Length	Vocabulary			
1	short	king/queen	boy/girl	maid/chef	clerk/judge
1	long	actress/bishop	soldier/sailor	mother/daughter	brother/sister
2	short	Dutch/Thai	Swiss/Czech	Greek/Basque	French/Welsh
2	long	Scottish/English	Russian/American	Spanish/Italian	Danish/Swedish
3	short	fat/thin	tall/short	strong/weak	young/old
3	long	hungry/thirsty	quiet/noisy	generous/modest	alert/innocent
4	short	rich/poor	good/bad	sane/mad	shy/gruff
4	long	graceful/awkward	peaceful/violent	gentle/heavy	healthy/delicate

Table D.1: The vocabulary used in Experiment X listed by Dimension (abbreviated as Dim.) and Word Length.

D.2 Results and analysis for Experiment X

Reaction Times

The reaction time data was trimmed to two standard deviations by subject. The outliers were evenly distributed across mode, matchtype, word length, subject, sentence, task and vocabulary item.

In order to test the general effects of task, predictability and word length a repeated measures ANOVA was conducted on the reaction time data. The fixed factors were task (normal—no secondary task, suppression and tapping), predictability (predictable or unpredictable text), word length (short or long text), individual (first or second introduced), property (1 to 4) and subjects was the only random factor.

There were significant main effects of predictability, word length, property and individual. There were four significant interactions: task by property, predictability by property, word length by property, property by individual and one three-way interaction between predictability, property and individual.

The interaction between task and predictability ($F(6,282) = 13.52, p < 0.0001$) showed that the SOE was affected by task and possibly showed that Suppression flattened the shape.

Sentences with short words were read faster than sentences with long words ($F(1,47) = 51.53, p < 0.0001$), 1.73 sec compared with 1.86 sec. The interaction between word length and property ($F(3,141) = 7.33, p < 0.0002$) showed that the size of the difference between short and long texts increased as more properties were read.

These results confirm our prediction that predictable texts are faster than unpredictable ones and that articulatory rehearsal is used. However, the lack of any interactions with task and word-length suggests that our subjects were not suppressing conscientiously.

The interaction between property, individual and property shows that in predictable modes the SOE is very similar for both individuals and that in unpredictable modes the SOE is different. This suggests that unpredictability has introduced asymmetry between the two individuals as predicted.

In order to investigate the effects of mode the predictable and unpredictable halves had to be analyzed separately because there were different numbers of modes in the two halves.

An ANOVA was conducted on the predictable data with word-length (short and long), task (normal, suppression and tapping), individual (first and second mentioned), property (one to four) and mode (I×I and P×P) as fixed factors and subjects (48 levels) as the only random factor.

There are main effects of property ($F(3,141) = 7.92, p < 0.0002$), word length ($F(1,47) = 38.32, p < 0.0001$) and individual ($F(1,47) = 5.97), p < 0.02$). Stenning, Shepherd and Levy (1988) found main effects of word length and property but no main effect of individual. In this experiment sentences about the second individual are read faster than the first individual (1.65 seconds and 1.73 seconds respectively).

There is only one interaction with word length which is with property ($F(3,141) = 3.24, p < 0.03$) and shows that the difference between short and long texts for properties three and four is about twice the size of the difference for properties one and two. The only interaction which Stenning, Shepherd and Levy (1988) found that included word length was between word length, mode, individual and property. They interpreted their interaction as an articulatory review which took place on the fourth property of the first individual in I×I texts because the difference they observed could be accounted for given Baddeley, Thomson and Buchanan's (1975) observed rate of articulatory rehearsal (0.2 seconds per syllable). In this experiment the difference between short and long conditions barely reaches the time taken to rehearse one syllable so the interaction is unlikely to be articulatory review.

Stenning, Shepherd and Levy (1988) found an interaction between individual, property and mode which is also present in this experiment ($F(3,141) = 20.36, p < 0.0001$), Figure D.1 shows the means. However, the two interactions seem to have a dissimilar nature. In this experiment the reading time patterns for I×I texts are very different from P×P texts. As the number of properties known of an individual increases so does the reading time in I×I

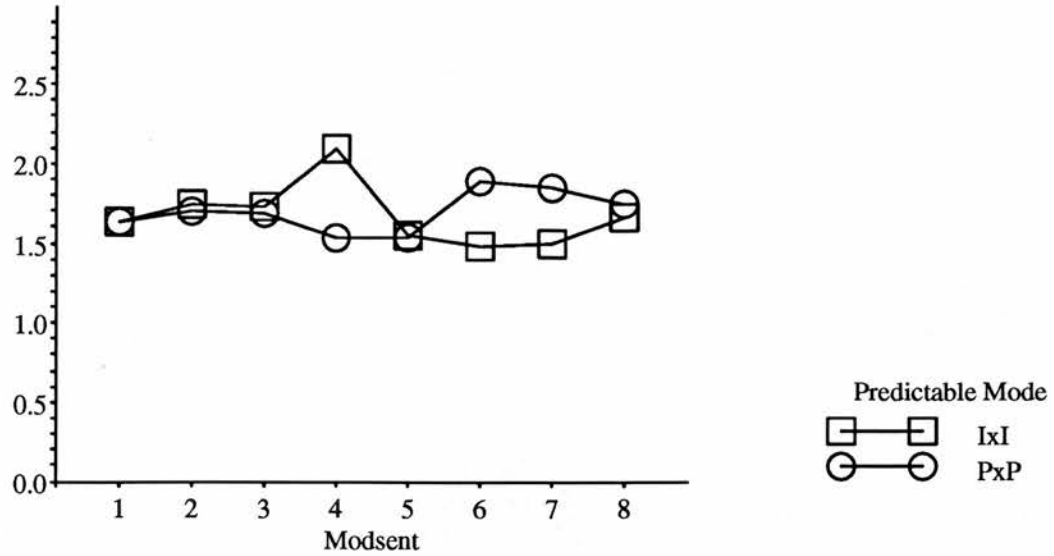

Figure D.1: Mean reading times (secs) as a function of Modsent and Mode.

texts but the reverse seems to be true of P×P texts which show a decrease after the second property.

There are three interactions which include task: task and property ($F(6,282) = 9.79, p < 0.0001$); task, individual and mode ($F(2,94) = 3.2, p < 0.05$); task, property and mode ($F(6,282) = 3.26, p < 0.005$). The interaction between task, mode and individual show that under suppression and tapping the difference between individuals for I×I texts is halved in comparison to the normal condition. It also shows that for P×P texts the first individual is faster than the second but in I×I texts the second individuals is faster. The interaction between task, property and mode shows no clear trend except that P×P texts seems to be affected by task more than I×I texts.

A similar ANOVA was conducted on the un-predictable data with task, word-length, individual, property and mode as the fixed factors with subjects as the only random factor. There were main effects of word-length ($F(1,47) = 27.76, p < 0.0001$), individual ($F(1,47) = 8.86, p < 0.005$), property ($F(3,141) = 22.96, p < 0.0001$) and mode ($F(7,329) = 9.97, p < 0.0001$). The effects of mode, individual and property all interacted in pairs and were all contained in the three-way interaction between interaction individual, property and mode ($F(21,987) = 18.18, p < 0.0001$). Figure D.2 shows the mean reaction times for each sentence by Mode and Modsent.

There was one interaction with task which was between task and property ($F(3,141) = 9.53, p < 0.0001$). It showed that in the secondary task conditions the rise in reading time with property was greatly reduced. There was also one interaction with word-length between word-length and property ($F(3,141) = 4.73, p < 0.004$) which shows the difference between short and long descriptions increases as the the number of properties increases.

A second ANOVA was conducted on the un-predictable data. The fixed factors were task, mode, status (primary or secondary), property and word-length and the single random fac-

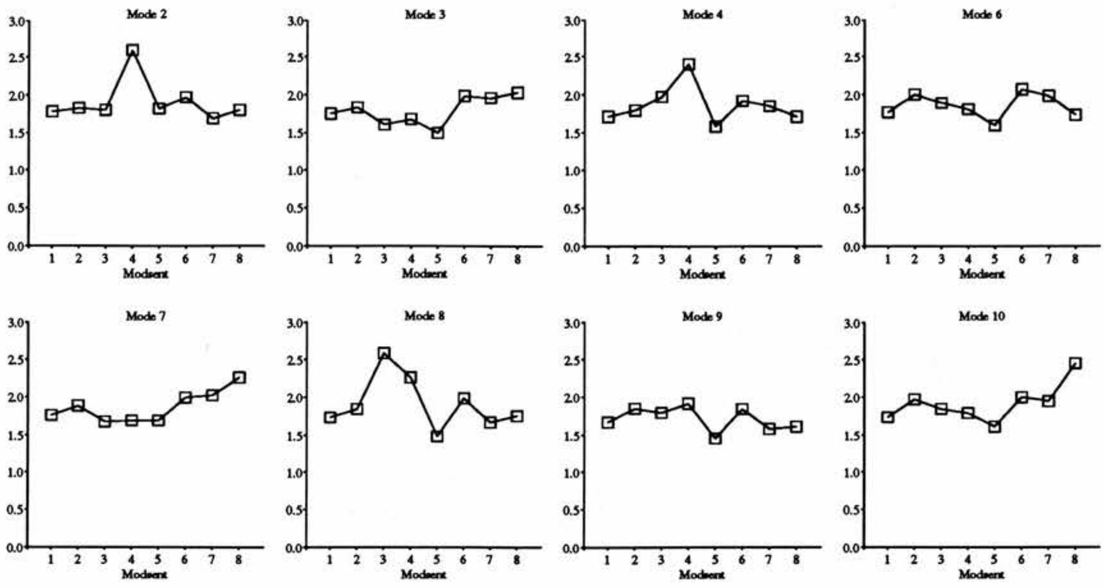

Figure D.2: Mean reaction times over subject by mode, individual and property

tor was subjects. There were main effects of mode ($F(7, 329) = 9.97, p < 0.0001$), status ($F(1, 47) = 56.67, p < 0.0001$), property ($F(3, 141) = 22.96, p < 0.0001$) and word-length ($F(1, 47) = 27.76, p < 0.0001$).

There was one interaction with word-length which was with property ($F(3, 141) = 4.73, p < 0.004$). This showed that the difference between short and long texts increased with property. Task interacted with property ($F(6, 282) = 9.54, p < 0.0001$) and showed that a secondary task removed the semantic ordinal effect although this was possibly more marked in the Suppression condition.

The rest of the interactions were all contained in the interaction between mode, status and property ($F(21, 987) = 11.25, p < 0.0001$). Figure D.3 presents the means over subject by mode status and property. It shows that the pattern of property reading times is different for the leading and trailing properties (or different by status) and that these patterns are affected by the pattern of reference which determines when there is a change of status or a change of reference. For the primary individual as more properties are learnt then the reaction time is longer and the large peaks are due to switches of reference. The secondary individual's reaction time peaks at the second property and the third and fourth properties are very similar.

Analyses of data collected in previous experiments (Experiments I and II) suggested that word length effects would be closely related to switching phenomena so an ANOVA was carried out on the unpredictable data to investigate this. In order to make the data sufficiently dense to reduce the effects of unbalanced cells subjects were grouped in threes at random. The fixed factors were contour (switch or continuation of reference), status (primary or secondary), word-length (short or long) and task (normal, suppression or tapping) and subjects was a 16 level random factor.

There were three main effects of contour ($F(1, 15) = 55.54, p < 0.0001$), status ($F(1, 15) = 171.71, p < 0.0001$) and of word-length ($F(1, 15) = 18.80, p < 0.0001$) short words were faster

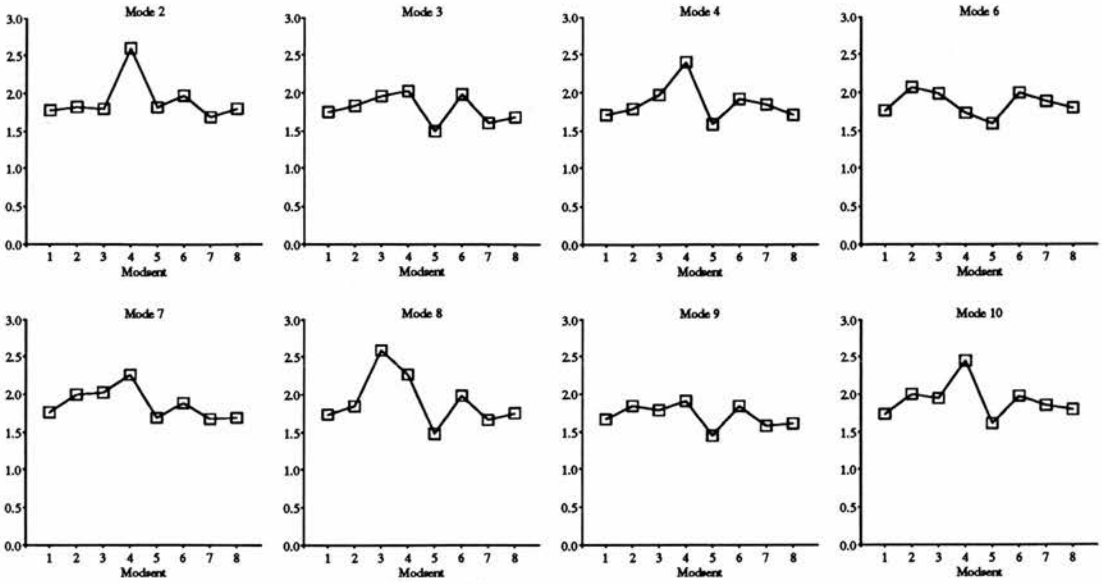

Figure D.3: Mean reaction times over subject by mode, status and property. Status and Propert and have been combined to form an analogue of Modsent where the first individual is replaced by the primary individual and the second individual by the secondary individual.

than long ones (means were 1.85 and 1.93 respectively).

There was an interaction between contour and status ($F(1,15) = 41.7, p < 0.0001$) and one between task, contour and status ($F(2,30) = 6.03, p < 0.007$). The three-way interaction showed that for the secondary individual under normal conditions a switch was read faster than a continuation but in the other two conditions there was little difference. The primary individual was unaffected by task. Figure D.4 shows the means for this interaction.

In order to assess the effects of a text's match structure an ANOVA was done on all the trailing properties on the second, third and fourth dimensions because these are the points in a text where match is defined. The ANOVA had task, word-length, property and match as fixed factors with subjects as a random factor. There are main effects of word length ($F(1,47) = 47.78, p < 0.0001$, short = 1.77, long = 1.87), match ($F(1,47) = 69.92, p < 0.0001$, mismatch = 1.89, match = 1.76) and property ($F(2,94) = 4.41, p < 0.02$, means from 2 to 3: 1.82, 1.78, 1.87).

There are three interactions with property: property by match ($F(2,94) = 11.28, p < 0.0001$), property by word length ($F(2,94) = 6.20, p < 0.004$) and property by task ($F(4,188) = 12.19, P < 0.03$). The interaction between property and match shows that the difference between matched and mismatched properties decreases as more properties are discovered. The interaction between word length and property shows that the difference between short and long texts goes from 0.07 to 0.13 to 0.11 seconds. The interaction between task and property shows that the V-shape (on the last two properties) which is found in normal and tapping conditions but not in the suppression condition where the usual rise on the last property is absent.

The other two interactions are between match and word length ($F(1,2) = 6.73, p < 0.02$) and between match and task ($F(2,94) = 3.8, p < 0.03$). The interaction between match and word

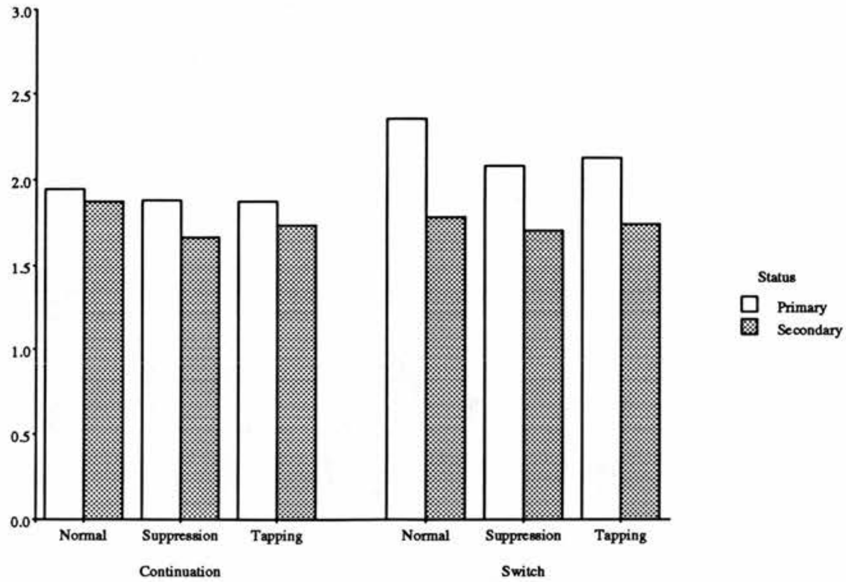

Figure D.4: Reaction times by contour, status and task

length shows that the difference between short and long texts is much smaller for matched properties than for mismatched one. The interaction between task and match shows that the difference between matched and mismatched properties varies with task, normal having the largest difference and tapping having the smallest.

Recall

An ANOVA was done on the cued scores with task, predictability and word-length as fixed factors with subjects as the only random factor. There were three significant main effects of task, word-length and predictability and no interactions. Scores were lower in the suppression and tapping conditions ($F(2,94) = 16.89, p < 0.0001$), 6.84 and 6.95 compared with 7.20. Scores were lower in un-predictable texts ($F(1,47) = 105.68, p < 0.0001$) than in predictable ones, 6.70 and 7.28 respectively. Short texts were more accurately recalled than long ones ($F(2,47) = 24.67, p < 0.0001$), 7.09 compared with 6.89.

A similar ANOVA was done on the bestfit scores. Again there were three significant main effects of task, predictability and word-length and no interactions. The recall was less accurate in secondary task conditions ($F(2,94) = 15.16, p < 0.0001$): the means were, 7.52 for normal, 7.30 for suppression and 7.34 for tapping. Predictable texts were recalled better than unpredictable ones, ($F(1,47) = 11.63, p < 0.002$), 7.44 compared with 7.33. Again short texts were more accurate than long texts ($F(1,47) = 53.49, p < 0.0001$), 7.49 compared with 7.29.

The set of unpredictable modes is made up of two groups (see Chapter 2), one of which causes more confusions of the order of introduction than the other. An ANOVA was done on the number of confusions per subject per task per mode group with mode and task as fixed factors with subjects as a random factor. There was a main effect of mode group ($F(1,47) = 78.88, p < 0.0001$), fewer confusions being made in mode group one (0.43) than in mode group two (2.14). The interaction between mode group and task was significant

	Modegroup 1				Modegroup 2			
Task/Mode	2	4	8	9	3	6	7	10
Nothing	5	9	5	4	15	20	32	21
Suppression	5	7	3	3	14	38	38	34
Tapping	6	7	4	4	19	26	25	26

Table D.2: The number of texts where the identity of the two individuals is confused (where the scoring criteria conflict) out of 96 texts per mode

	Single error		Multiple error	
R-Individual 1				
Stimulus position:	First	Second	First	Second
	9.37	10.32	1.95	2.69
R-Individual 2				
Stimulus position:	First	Second	First	Second
	10.72	11.80	1.80	2.71

Table D.3: Percentage of single and multiple error as a function of stimulus position and order of recall (N=4608).

($F(2,94) = 3.34, p < 0.05$). For mode group 1, the means for normal, suppression and tapping were 0.48, 0.37 and 0.44 and for mode group 2 the means were 1.833, 2.58 and 2.00. The means suggest the suppression causes more confusions than normal or tapping but only for mode group 2 modes. Table D.2 shows the total number of confusions by mode and task.

An ANOVA was done on the bestfit scores for the texts where there was a conflict and where the cue had been followed. The difference between 7.20 (where the cue was not followed) and 7.33 (where the cue was followed) was not significant ($F(1,47) = 2.93, p = 0.09$). This shows that when subjects confused the order of introduction of the individuals they were still able to bind the properties together successfully.

Table D.3 shows that the second recalled individual is less accurate than the first recalled individual in all but one case. It also shows that the individual introduced second is less well recalled then the first individual. However, the differences between the accuracies are much smaller than the differences found in Stenning, Shepherd and Levy (1988). Therefore in this experiment it seems that subjects were less prone to interference between individuals.

Table D.4 shows the percentage of errors by property across matchtype. It is clear that for each error category more errors occur if the dimension is mismatched. There is a tendency for more errors to occur on the first recalled individual. If a dimension mismatches then there are usually more double errors (an error is made on both individuals) but if it matches then the number of double errors is smaller than the other types. The third dimensions shows an

	Property 1		Property 2		Property 3		Property 4	
	+	−	+	−	+	−	+	−
1st object	–	0.72	2.47	4.90	5.95	9.98	4.82	5.34
2nd object	–	1.35	3.17	4.21	4.99	5.60	4.43	5.08
Both	–	2.06	2.21	6.30	2.26	6.51	2.04	5.47

Table D.4: Percentage of first, second and both object errors as a function of matched and mismatched properties. Object order is defined by recall order.

		Single errors			Multiple errors		
		No	Supp	Tapp	No	Supp	Tapp
pred	R-Ind 1	15.75	19.27	20.07	4.03	4.56	4.29
	R-Ind 2	16.53	21.87	21.87	2.73	4.95	3.78
u-pred	R-Ind 1	17.71	22.53	22.27	2.47	6.51	5.99
	R-Ind 2	20.31	31.90	22.66	3.64	4.82	7.16

Table D.5: Percentage of texts with single and multiple errors as a function of predictability and task. (N=768)

			R-Individual 1		R-Individual 2	
Predictability	Task	Word-length	Single	Multiple	Single	Multiple
Predictable	Normal	Short	12.8	3.4	12.2	2.6
		Long	18.8	4.7	20.8	2.9
	Suppression	Short	19.0	2.9	20.6	3.9
		Long	19.5	6.3	23.2	6.0
	Tapping	Short	18.2	4.4	20.1	2.3
		Long	23.2	4.2	23.7	5.2
Upredictbale	Normal	Short	12.8	1.6	15.4	3.1
		Long	22.7	3.4	25.3	4.2
	Suppression	Short	17.2	6.6	27.1	4.9
		Long	27.9	6.5	36.7	4.7
	Tapping	Short	15.9	5.7	20.6	4.9
		Long	28.6	6.3	24.7	9.4

Table D.6: Percentages of errors by individual, word length, task and predictability (N=384)

unusually high error rate on the first individual when the dimension mismatches. It looks as if when a dimension is mismatched subjects will remember that fact but may become confused about which way round the properties go—this explains why double errors are high when a dimension mismatches. However, when a dimension matches subjects often forget this and suppose that the dimension mismatches otherwise the number of double errors would be higher. The third dimension shows unusual behavior because the error rate on the first individual when the dimension mismatches seems to show that subjects forget that the dimension mismatches and assume that the first individual has the same property as the second individual.

Table D.5 shows that task appears to interact with predictability. The percentage of single and multiple errors is higher for both individuals under conditions of suppression and tapping which agrees with the first two ANOVAs described above. However this table also shows that the percentage of single errors is unusually high for the second individual under suppression conditions for unpredictable texts.

Table D.6 shows that although the percentage of single errors made on the second individual in unpredictable texts under suppression conditions is particularly high the difference between short and long texts is unaffected. For unpredictable texts it looks as if the difference between multiple error frequencies for short and long texts may be abolished by suppression for both individuals. Two ANOVAs were conducted one for frequency of multiple errors on the first individual and for frequency of multiple errors on the second individual. The factors were word-length, task, predictability with subjects as a random factor. Both ANOVAs showed a significant effect of task and the ANOVA for the frequency of errors on the second individual showed an effect of word-length and word-length as well. However neither ANOVA showed an

		Single errors							
		R-Individual 1				R-Individual 2			
Task	Word-length	1	2	3	4	1	2	3	4
Normal	Short	1.30	3.12	4.29	4.03	0.65	3.12	5.98	4.03
	Long	3.25	5.59	6.90	4.94	2.73	6.25	10.15	3.90
Suppression	Short	1.30	5.20	6.77	4.81	1.30	5.46	9.11	7.94
	Long	3.12	5.98	7.29	7.29	2.86	7.29	11.84	7.94
Tapping	Short	1.17	4.42	5.59	5.85	1.82	4.42	8.98	5.07
	Long	3.12	6.77	8.59	7.42	2.21	6.11	8.85	7.03
		Multiple errors							
Normal	Short	0.65	1.56	1.69	1.43	0.65	2.08	1.95	1.56
	Long	1.30	2.47	2.34	2.21	0.52	2.08	2.34	2.34
Suppression	Short	1.17	2.73	3.38	2.86	0.78	2.86	3.51	2.34
	Long	1.56	3.64	4.42	3.90	1.17	2.21	3.90	3.51
Tapping	Short	1.43	2.99	2.99	3.12	0.52	2.21	2.08	2.47
	Long	1.04	3.12	3.77	3.12	1.43	3.51	5.33	4.81

Table D.7: Percentages of single and multiple errors by properties within individuals by task and word-length. The number of texts in each combination of task and word-length was 768

	Dimension			
	1	2	3	4
Single errors				
R-Individual 1	2.21	5.19	6.58	5.73
R-Individual 2	1.93	5.45	9.16	5.99
Multiple errors				
R-Individual 1	1.19	2.76	3.10	2.78
R-Individual 2	0.85	2.50	3.19	2.84

Table D.8: Percentages of multiple and single errors across properties within individuals (N = 4608).

interaction between task and word-length.

The number of errors of both types are unaffected by task and word-length across properties which Table D.7 shows.

Table D.8 shows that single errors are the most common type of error on both individuals. Stenning, Shepherd and Levy (1988) found a similar result. It is noticeable that the most single and multiple errors occur on dimension 3. This difference is largest for single errors on the second recalled individual. Stenning, Shepherd and Levy (1988) found the only case where the third dimension was the least accurate was in the same position: single errors on the third dimension of the second recalled individual.

The errors for a single text were classified using Stenning, Shepherd and Levy's (1988) scheme and is repeated in Appendix D.4. The suppression task clearly affected the accuracy of the second recalled individual in unpredictable texts and classifying them showed what the particular errors were. Table D.9 shows that in unpredictable texts single errors on matched and mismatched dimensions are affected by the suppression task and that individual polarity errors are also affected though less so. Predictable texts show very little evidence of a similar pattern.

Predictability	Error category	Task			
		Normal	Supp	Tapp	Total
Unpredictable	sg2+	55	66	46	167
	sg2-	20	55	27	102
	ipol	45	56	47	148
	is1+	7	9	8	24
	is1-	1	8	2	11
	2cdf	5	14	12	31
	dhdf	8	10	6	24
	ppol	8	20	13	41
	pp+s	2	2	1	5
	mirr	0	0	1	1
	misc	2	4	8	14
	Total	153	244	171	568
Predictable	sg2+	20	27	26	73
	sg2-	32	34	33	99
	ipol	42	55	64	161
	is1+	1	6	5	12
	is1-	2	4	6	12
	2cdf	7	8	6	21
	dhdf	4	7	1	12
	ppol	12	24	19	55
	pp+s	2	2	2	6
	misc	1	1	2	4
	Total	123	168	164	455

Table D.9: Number of texts with a single error on the second recalled individual as a function of task and predictability

D.3 Discussion

The reading time analyses will be discussed then the recall analyses.

D.3.1 Reading times

The ANOVA's showed that word length had a main effect which suggest articulatory involvement. Although there were several interactions with task there were no interactions which included an interaction between task and word length as predicted which means that the effects of word length were not genuine articulatory effects or simply that suppression was not "working". However, word length did interact with property and showed that the size of difference between short and long texts increased as more properties were discovered which is consistent with rehearsal of the referenced individual's description. Task also interacted with property and seemed to show that suppression flattened the usual increasing function. These results suggests that subjects were rehearsing the referenced individuals description and that suppression was reducing this rehearsal.

There was an interaction between task, contour and status which suggested that the primary individual was unaffected by task whereas the secondary individual was which is consistent with the supposition that the secondary is more dependent that the primary individual on rehearsal. However, it was unclear whether suppression had a different effect from tapping.

Therefore the result may simply be one of attentional disruption. Task also interacted with match but Tapping seemed to have a different effect from Suppression and Normal so it was hard to explain how task related to match processing.

Word length effects were reliable and interacted with several other variables. Because word length did not interact with task they cannot be attributed to articulatory rehearsal with certainty but they are suggestive. For example word length interacted with match and seemed to suggest that when properties matched subjects did not rehearse both properties. This seems a reasonable strategy to adopt if both descriptions are being rehearsed.

The rest of the effects in the analyses either show an attentional effect on some interaction or they replicate results from previous MIT experiments. For example, the ANOVA on the unpredictable data contained an interaction between mode, property and individual which had been found in Experiment II and showed that the pattern of reference in a text affected the sentence reading times.

Overall the reading times analyses failed to show that the observed word length effects were articulatory because they did not interact with task. The word lengths did occur in places which gave a consistent interpretation but were not conclusive. Task did have various effects and a few interactions but in many cases Suppression appeared to be indistinguishable from Tapping and therefore the effects of Task were mainly attentional. In spite of Task's disappointing effects there were some interactions with property which hinted that suppression was having a different effect from tapping and normal.

D.3.2 Recall results

The general pattern of confusions in the unpredictable modes again replicated the pattern shown in Experiment I for cued recall and provide further evidence for the distinction between primary and secondary individuals. The first two ANOVAs on the bestfit and cued scores showed a similar pattern which agreed with the reading time results. Short texts were easier than long texts, predictable texts were easier than unpredictable texts and the secondary tasks contributed an equal load which reduced the recall accuracy.

The interaction between modegroup and task supports a very interesting pattern of results in the data. Modegroup 1 modes which do not contain a change of status show very little effect of Task. However, the Modegroup 2 modes in which there is a change of status do show an effect of Task. The effect of the secondary tasks are not the same: it appears as if it is suppression alone that causes an increase in the number of confusions in Modegroup 2 modes. This result seems very close to the expected results. If suppression is really blocking articulatory rehearsal then it is showing that articulatory rehearsal is used to maintain identity information which is particularly crucial after a change of status. This agrees very closely with the observation in Experiment IX's data that the number of dual segment commentaries increases after there has been a change of status.

The error analyses generally show that suppression and tapping have a similar effect. However, there is some evidence that suppression does increase the number of single errors on the second recalled individual in unpredictable texts. Unfortunately this effect does not interact with word length and is unrelated to particular Properties although dimension 3 did show an unusually high error rate for individual 1 mismatches. Single and multiple errors are constant across Properties.

Predictable and Unpredictable texts show a different pattern when classified by Stenning, Shepherd & Levy (1988)'s scheme. Unpredictable texts show an effect of suppression on single errors which predominantly occur on a mismatched dimension on the second individual. Presumably suppression is causing an increase in errors on the mismatched dimension because subjects are no longer able to use articulatory rehearsal. Again, there are no effects of word length so the evidence is not complete.

The results from both reading time and recall data both suggest that suppression was having more than attentional effect because there were instances where tapping had a similar effect to normal conditions which were different from suppression conditions. What is more suppression is related to changes of status and confusions of identity which is what the data from Experiment IX indicated. The regression models from Chapter 2 showed that articulatory rehearsal was possibly more important for the secondary individual which is usually recalled second and there was some evidence that the second recalled individual was indeed disrupted by suppression.

D.4 Error-type classification

Consider two individuals which have been described in a MIT text: a tall, happy, Swiss dentist and a short, happy, French nun.

Single property errors are the simplest error and are a misrecalled property of an individual which can be of either individual and either matched or mismatched. For example, sg2+ refers to a single error on the individual recalled second on a matched dimension and sg2- corresponds to a single error on the individual recalled second on a mismatched dimension.

If two properties are incorrectly recalled then the error can be classified in several ways. If the two errors occur on the same dimension then the error is called a polarity error. If the dimension is mismatched then the error is called an individual polarity error and if the dimension is mismatched then it is called a property polarity error.

If the two errors occur on different dimensions then the error may be homogeneous because both dimensions have the same match status or the error may be complementary because the dimensions have different matching status. The two errors may occur on the same individual or on different ones.

A mirror error is where the match structure of all the dimensions except the introducer are reversed.

Table D.10 shows examples of the errors discussed.

	Match structure					
Indidivual	–	–	+	–	Response type	Abbreviation
R-1	dentist	Swiss	happy	tall	Correct	corr
R-2	nun	French	happy	short		
R-1	dentist	Swiss	happy	tall	Single error on R-2 matched	sg2+
R-2	nun	French	sad	short		
R-1	dentist	Swiss	happy	tall	Single error on R-2 mismatched	sg2-
R-2	nun	French	happy	tall		
R-1	dentist	Swiss	happy	short	Individual polarity error	ipol
R-2	nun	French	happy	tall		
R-1	dentist	French	happy	short	Indidivual polarity with "sg1-"	is1-
R-2	nun	French	happy	tall		
R-1	dentist	Swiss	sad	short	Individual polarity with "sg1+"	is1+
R-2	nun	French	happy	tall		
R-1	dentist	French	happy	tall	Double complementary on R-1 and R-2	2cdf
R-2	nun	French	sad	short		
R-1	dentist	French	happy	short	Double homogeneous on R-1 and R-2	dhdf
R-2	nun	French	happy	tall		
R-1	dentist	Swiss	sad	tall	Property polarity error	ppol
R-2	nun	French	sad	short		
R-1	dentist	French	sad	tall	Property polarity with single	ppol+s
R-2	nun	French	sad	short		
R-1	dentist	Swiss	happy	short	Mirror image matching structure	mirr
R-2	nun	Swiss	sad	short		

Table D.10: Examples of error categories and abbreviations.

Appendix E

Details of the author's contribution to the work reported herein

Chapter 2 contains descriptions of two experiments, Experiment I and Experiment II. Experiment I was designed and run by Mukesh Patel who reported an analysis in his unpublished PhD thesis (Centre for Cognitive Science, University of Edinburgh, 1990) and in Stenning, Patel and Levy (1987). I carried out a different analysis which depended on his earlier work and is reported in Chapter 2 and described as the pilot experiment reported in Stenning, Nelson, Levy, Patel and Gemmell (in press) which is reprinted in Appendix F. I designed (jointly with Keith Stenning, Martin Gemmell and Joe Levy), ran (jointly with Martin Gemmell) and analysed (jointly with Keith Stenning) Experiment II which is also reported in Stenning *et al.* (in press).

Of the experiments reported in Chapter 3, I designed each experiment in collaboration with Rosemary Stevenson and Keith Stenning. I also generated all the materials and analysed all the data from the experiments which are reported.

Chapter 4 contains the description of just one experiment which was designed by Mukesh Patel and run by Morag Brown and reported in her undergraduate dissertation (Edinburgh University, unpublished). I logged the data from it (jointly with Martin Gemmell) and produced all the analyses which are reported here and in Stenning *et al.* (in press) as Experiment 2.

In Chapter 5 I designed all the experiments in collaboration with Joe Levy and Keith Stenning. I also generated all the materials and ran all the experiments with Joe Levy. I produced all the analyses which are reported.

Appendix F

Reprint of paper containing research reported herein

Representations of individuals and the processing of reference change

Keith Stenning[1] Alexander W.R. Nelson[1]
Joe Levy[1] Mukesh J. Patel[2] Martin Gemmell[1]

[1]Human Communication Research Centre
University of Edinburgh
2 Buccleuch Place
Edinburgh EH8 9LW

[2]School of Computing and Cognitive Sciences
University of Sussex
Falmer
Brighton BN1 9QH

Running title: Reference change

Abstract

Stenning, Shepherd and Levy (1988) showed that when simple texts switch reference predictably between individuals, changes of reference neither cost reading time nor degrade memory performance. The present experiments examine the effects of unpredictable referential change. Experiment 1 demonstrates that unpredictable reference change does cost processing time, as a function of the amount known about the referent to which attention shifts. Analysis reveals a distinction between *primary* and *secondary* individuals related to referential change. It also reveals word length effects, both decelerations and accelerations proportional to description length, which are interpreted in terms of use of the articulatory loop (Baddeley, 1986). Experiment 1 reveals involvement of primary/secondary status in the process of switching reference, and shows that the word length effects cannot be interpreted in terms of frequency. Experiment 2 strengthens support for the primary/secondary distinction and confirms the use of the articulatory loop. The present results suggest a central role for distributed information about sequence in representing complex semantic structures both in immediate and in long term memory. Predictable switching costs no time because the transparency of the relation between surface sequence and underlying semantic structure is preserved. The distinction between primary and secondary individuals emerges with unpredictable reference because it restores this transparency.

Acknowledgements

The support of the Economic and Social Research Council UK (ESRC) is gratefully acknowledged. The work was part of the ESRC funded Human Communication Research Centre (HCRC). Thanks to Morag Brown for her assistance with Experiment 2. The authors would like to thank their anonymous reviewers for helpful suggestions.

Introduction

Discourses introduce individuals and attribute properties and relations to them. Several references may be made to an individual, each time attributing new properties. These references may either be made successively or with references to other individuals interspersed between. The structures and processes involved in introducing referents, reidentifying them and adding to their representations must form the core of a theory of discourse processing. The strong effects of different orderings of material are intuitively evident to the language user and have been systematically studied by rhetoricians.

Work on anaphora resolution, verbal reasoning and on memory itself has been concerned with issues of ordering references in text and explaining the effects on reading through theories about the processor, as we will discuss below. But in all these literatures, the *representation* of the binding of attributes to individuals is assumed to need no analysis or explanation. Consider the following problem: after reading a description of a short fat Polish person and a rich clever teacher, how do we keep these properties in the correct configuration and not recall reading about a fat teacher, or a rich Polish person, or ... . Links between representations of individuals and representations of properties are assumed to be an adequate account of binding. This is despite the fact that human memory experiences great difficulty in maintaining patterns of binding when many plausible alternatives are available, and does so as a function of the *content* of what is to be linked. Anaphor resolutions take place within representations of bindings accumulated prior to the anaphor, and lead to the construction of representations of new bindings. Pragmatic information not only plays a role in the *choice* between alternative antecedents, but also in the *construction* of representations of the outcomes of these choices. Similarly, verbal reasoning requires establishing working memory bindings of properties and relations to novel individuals. This sub-task is a major component of the reasoning task and is strongly affected by content (e.g., Johnson-Laird, 1975).

Studies of the part played in the selection of antecedents by syntactic, semantic, thematic and pragmatic factors describe what patterns of pronominalisation languages adopt and are a valuable source of clues as to what the processor's memory is like, but there is a deeper level at which a computational *explanation* is required. Why do language users adopt these patterns? For example, many texts, particularly narratives, are organised so that one individual serves as a continuing focus of attention throughout (the thematic subject), and certainly there is usually one such focus within any particular episode (Garrod & Sanford 1988). This individual is more accessible and is treated differently by the processes which find antecedents for pronouns. Is this because narrative explores character, and does so by analysing one character in relation to others, or is it because our text processor has a computational architecture for which there is an advantage in focussing on one referent throughout a text? Only a computationally explicit theory of working memory as it is deployed in text processing will provide the answer.

The literature on verbal inference does provide deeper experimental analysis of the effects of unpredictability of structure on reading (see McGonigle & Chalmers, 1986 for a review). Several authors have used self-paced reading time to investigate their subjects' construction of representations whilst reading texts (e.g. Ehrlich & Johnson-Laird, 1982; Kieras & Just, 1984). They have generally concluded that introducing a new referent slows reading, and that switching reference from one individual to another already known, also results in longer reading times. These effects are interpreted by Ehrlich and Johnson-Laird as effects due to the increased memory load of having to hold temporarily unresolved information.

Stenning and Levy (1988) proposed a model of attribute binding based on distributed representations. The theory is distinctively two-level: whole high-level patterns of binding of properties to individuals are the result of inferences from fragmented low-level data. Stenning and Levy's model of binding is built on their earlier experimental data from subjects remembering pairs of individuals each with four properties. This data consisted of the frequencies with which various patterns of error occurred in recalling the pairs of individuals. Errors can distort between one and four properties. The types of errors observed showed that the underlying representation is redundant and its fragments are logically related—too many complex errors and too few simple ones occur to be explained on the basis of single link encodings of binding. In the model, subjects' knowledge that a certain individual has a property is distributed over about fifteen simple existential propositions about the pair of individuals described. This fragmented representation is 'synthesised' into a pattern of binding for the whole ensemble by a PDP system operating as a constraint satisfaction mechanism. When noise is injected in the form of randomly changed truth values of the primitive propositions, the pattern of errors made by the network accords well with the pattern in subjects' data. In this system, binding only emerges as the result of an inferential process operating on the lowest level of representation. Binding in the full sense of binding of textual attribute to textual individual is a high level achievement, though there are bindings between items of low level information in the low level fragments of representation. An account of the content sensitivity of memory is still required, but we claim that it can naturally be attached at this lower level.

In Stenning & Levy's model, distributed representation has a rather precise definition because the meanings of the units are defined in logical terms. At the lower level, there are units whose activations represent the truth values of propositions such as 'There is a happy Polish person', 'There are two people of the same nationality', 'There is a tall happy German person', 'There is someone who is Polish and short ...'. These units represent their propositions *locally*. However, the total information about the bindings of the four property dimensions is not represented by any one of these units—it is *distributed* over them, and in a redundant (and possibly inconsistent) fashion. The network then performs a constraint satisfaction inference and outputs, at the higher level, a canonical description of the bindings which is again made up of local representations—ones from which individual's properties can be directly read off.

This explanation of binding offers a natural explanation of some effects of ordering in text which other theories do not. For example, spreading activation theories of immediate memory (e.g. Anderson, 1983) interpret the cost of changing reference as the cost of re-building activation. Repeated reference to an individual injects activation at its node in a semantic network. Activation spreads outward from this node and dissipates. Shifting reference to an individual represented by another node temporarily depresses the achieved level of activation. The time to restore a given activation level at the new referent's node is an increasing function of the node's 'fan', because activation will disperse more rapidly from a node with many outgoing arcs. However, this bottom-up explanation of the cost of referential change cannot explain Stenning, Shepherd and Levy's finding that predictable change costs no time. Here again, there is a close relation between the view taken of referential change and the mechanism of binding attributes to individual identities.

In Anderson's (1983) ACT*, single nodes in a semantic network represent individuals and are linked to nodes representing their attributes by content independent arcs. The network is passive with regard to processes which establish consistency — only an external inference engine working over the network can do that. In contrast, in Stenning, Shepherd and Levy's (1988) model of binding, the binding of attributes to individuals is only represented by the outcome of an active process of resolving a pattern of binding which best fits a distributed set of propositions about an array of individuals. This makes remembering bindings a high-level achievement applying to larger discourse structures, rather than a low-level representation of piece-meal links. It therefore opens up the possibility of explaining why holistic properties of text such as predictability of reference should have repercussions on the memory structures that result, and why, as we will show, *predictability* rather than *continuity* of reference is the determinant of processing.

Representing binding is not the only task in processing text, nor is the memory model proposed by Stenning and Levy intended to exhaust all aspects of working memory. It is a model of semantically interpreted material and does not consider, for example initial phonological encodings. This issue arises because some of the observations of Stenning, Shepherd and Levy and of the present studies might be explained by articulatory rehearsal effects. Moreover, we conclude that sequence plays an important role in the representation of binding. Given the presence of articulatory/phonological effects and their known sequence-preserving properties, they are clearly relevant to any theory of binding.

Baddeley's (see Baddeley, 1986, for a review) theory of Working Memory has been the source of much of the work on articulatory rehearsal processes. Baddeley argues that working memory is a set of slave systems each with specific coding characteristics organised around a central executive (and an interface to long term memory). Of these systems, the articulatory rehearsal loop (ARL) and its associated articulatory/acoustic store (AAS) will particularly concern us here.

Baddeley's theory proposes that there will be different sorts of information about antecedents held in different stores at different points in discourse processing, so it offers the possibility of explaining differential accessibility of material to subsequent anaphors. Although there is some evidence, chiefly from clinical studies, that the ARL/AAS is implicated in text processing (Vallar & Baddeley, 1984, but see Campbell & Butterworth, 1985 for counterargument), the evidence on which the working memory framework is presently based is chiefly from list memory tasks. These materials do not have the structure required to ask questions about representations of individuals or of processes involved in shifts of reference from one individual to another. For this reason, this framework has not yet developed an account of the interface between the slave systems of working memory and long term memory.

The highly constrained texts of Stenning, Shepherd and Levy's memory for individuals task (MIT) allow analysis of the rehearsal loop in text processing and thus provide a link between the memory literature and that on language processing. Stenning, Shepherd and Levy investigated ARL involvement in their task because rehearsal had been suggested as an explanation of the slowing of reading with increasing knowledge of referents — longer descriptions held in memory simply take longer to rehearse. They showed that there were word-length effects diagnostic of ARL involvement, but that they occurred only at isolated points in the regular texts used, and could not explain the relation between increasing knowledge of a referent and reading times for references to it.

The work reported here stems from Stenning, Shepherd and Levy's (1988) model of binding and their finding that, in simple texts describing two individuals in predictable sequences, the binding of attributes to individuals has a distributed representation and that *predictable* switching of reference between individuals costs no processing time or errors. The present experiments make the minimal changes necessary to extend Stenning, Shepherd and Levy's observations to texts with unpredictable reference. Their subjects read texts describing a pair of individuals, each with four properties, in one of two orders: either all the properties of the first individual were exhausted before the second individual was mentioned, or both individuals were described on the first adjective dimension (say profession), and then on the second (say nationality), and so on. The first order was called 'individual by individual' (I $\times$ I) and the second 'property by property' (P $\times$ P). An example of an I $\times$ I text would be: "There is a bishop. The bishop is Polish. The bishop is fat. The bishop is short. There is a dentist. The dentist is Swiss. The dentist is thin. The dentist is tall". The same description as a P $\times$ P text would be: "There is a bishop. There is a dentist. The bishop is Polish. The dentist is Swiss. The bishop is fat. The dentist is thin. The bishop is short. The dentist is tall." When only these two orders are used, reference is entirely predictable after the second sentence of any text. These predictable orders are of more than purely experimental interest: they are precisely the two orders which classical rhetoric prescribes for visiting a pair of topics (e.g. see Yates, 1966, on Cicero).

The plan of the paper is as follows. We briefly describe a pilot experiment where reference is made unpredictable by introducing seven different possible orders of attribution of properties to individuals. In the process of analysing the pilot data we discovered differences between the processes of switching reference to the two individuals, and complex word-length effects indicative of the use of the ARL/AAS system.

The observed asymmetries between the two individuals described were unexpected, as were the particular effects of word-length. Since the pilot was not specifically designed to analyse either these asymmetries or articulatory rehearsal, we then designed Experiment 1 which extends observations to new reference orderings and introduces control of frequency. The data from this experiment is then modeled. The model confirms the asymmetric word-length effects observed in the pilot and licenses two possible explanations for one of them. We designed Experiment 2 to investigate the possibilities and found clear evidence for a single interpretation. Experiment 2 was also designed to test predictions about the new reference orderings.

Finally, we interpret the theoretical consequences of these results for the representation of individuals, for reference change, and for the role of articulatory rehearsal in these processes. We propose that the binding of properties to individuals is distributed over representations of sequence, among other things, and that the time costs of changing reference are not due to the shifting of attention but to the extra construction of representations of disturbed sequence.

Where sequence is not disturbed, there are no time costs of switching reference.

Experiment 1

A pilot experiment was performed (see Stenning, Patel & Levy, 1987; Nelson, forthcoming) where reference was unpredictable in contrast to the predictable orders of reference used by Stenning, Shepherd and Levy. This was achieved by using seven different orders of reference (*modes*) instead of only two. Reading times for each sentence were recorded as well as errors made in recall. During the recall phase subjects were cued to recall the described individuals in order of their introduction or reverse order. Analysis of the recall errors showed that subjects confused the order of introduction more often in some modes than others. These confusions occurred most often for modes where a large proportion of the properties from the second introduced individual were described before the description of the first individual had become established. The confusions suggested an asymmetry in the treatment of the two individuals. This was also reflected in differences in reading time effects when reference changed. Early statistical modeling of the reading time data revealed word length effects. However, word length and frequency were confounded in the vocabulary used so a new experiment was designed to avoid this difficulty and provide more extensive data for reading time modeling.

Several new modes were added to those used in the pilot experiment to investigate some additional combinations of regression variables. The ten modes used are shown in Table 1.

(TABLE 1 ABOUT HERE)

A subsidiary design aim was to differentiate as far as possible between the effects of number of properties and syllables in descriptions, and word frequencies. For regression analysis it was desirable to decrease the correlation between numbers of properties and syllables in descriptions. It is not possible to decrease this correlation sufficiently without resort to extremely long words which make for unnatural descriptions. We opted to control frequency while reducing the correlation between numbers of syllables and properties as far as is possible with natural descriptions. Discriminating between syllable and property effects is the focus of another experiment reported elsewhere (Levy, Nelson & Stenning, in prep.) which additionally uses articulatory suppression techniques.

The vocabulary was arranged so that opposed pairs of properties come in LONG—LONG, LONG—SHORT and SHORT—SHORT pairs, and that this pattern is repeated in a high frequency and a low frequency vocabulary. Each text is then either entirely composed of high or entirely of low frequency words. The effects of frequency are then controlled, whilst correlation between numbers of syllables and properties is reduced as far as possible to aid regression modeling.

Method

Materials

The 48 words of the vocabulary came in six pairs for each of four property dimensions: profession, nationality, temperament and stature. Of each six pairs, one LONG—LONG, one LONG—SHORT and one SHORT—SHORT pair were high frequency words, and the same for low frequency words. The resulting vocabulary appears in Appendix .

High frequency words were defined as occurring in 20[1] or more of the 500 samples of 1,014,000 words in the Francis and Kucera (1982) word count, whereas low frequency words occurred in 10 or less such samples. Short words were of one syllable: long words were of more than one syllable (either two or three).

The description of a pair of individuals was generated by randomly choosing four contrasting pairs of attributes, one pair from each dimension. The *introducer* property dimension (profession) is always mismatched (i.e. the individuals differ). Individuals were matched (i.e. had identical properties) or mismatched on the other three properties equally often. Thus there are eight different structures of matching, referred to as *matchtypes* and annotated: $+++$, $++-$, $+-+$, $+--$, $-++$, $-+-$, $--+$, and $---$, where a '+' denotes a match and a '−' a mismatch, and the three symbols represent the non-introducing dimensions in temporal order of their attribution to the first individual. The order of the eight sentences within a description is determined by the mode.

Design

Frequency and mode were balanced: matchtype was randomised. Order of recall was not cued as it had been in the pilot experiment, simply to reduce the size of the design. Paragraphs of different frequencies of vocabulary, and of different modes occurred in random order.

Procedure

Thirty one psychology undergraduates participated as a course requirement. They were presented with the texts on a microcomputer monitor.

Subjects were provided with written instructions which were supplemented with detailed verbal instructions during the trial session. Subjects were instructed to read the texts as quickly as possible, consistent with recalling them accurately.

Each text was preceded by a *setting* which displayed the paired attributes from which descriptions were generated. Below is an example of a setting:

(bishop/dentist) (Swiss/Polish) (fat/thin) (short/tall)

The setting remained visible until the subject pressed the space bar. Pressing removed the current display and presented the next sentence. Times were measured between bar presses in centiseconds. At the end of each text, in order to disrupt superficial memory the subject was required to answer a simple question such as "Was there a fat bishop?" The response to the question was followed by a request to recall the individuals described by the most recent text, using a menu selection system. A menu then appeared offering a choice between the two contrasted properties on each dimension. After recalling one individual by making selections from the menu, the process was repeated for the other individual. Subjects were free to recall individuals in either introduced or reverse order. After the recall stage, feedback was given in the form of a single sentence description of both individuals presented (e.g. "There was a tall thin Swiss dentist and a short fat Polish bishop"). Subjects were provided with no other feedback on the accuracy of their recall of each individual. Subjects pressed the RETURN key to begin the next text presentation.

Results

Mean reading times for each sentence of each mode are shown in Table 2.

(TABLE 2 ABOUT HERE)

Reading Time Modeling

The development of the regression model of the reading time data from Experiment 1 investigated the same range of variables as the development of the regression model for the pilot data (Nelson, forthcoming).

The pilot experiment demonstrated an asymmetry between individuals that was not contingent on their order of introduction. In that experiment recall was cued but subjects often recalled the second introduced individual when asked to recall the first and *vice versa*. In these cases the *status* of the second introduced individual becomes similar to that of the first individual in the other modes as evidenced by the distinctive pattern of reading times when reference changes. The reading times showed a clear asymmetry between individuals of a *primary* and *secondary* status. These two sources of evidence lead to the following definition of primary/secondary status.

1. Assign the first individual introduced PRIMARY status.
2. If B2 (the second property of the second introduced individual) arrives before A2 (the second property of the first introduced entity), switch assignment of PRIMARY/SECONDARY.
3. If B3 precedes A3 (but A2 precedes B2), switch assignment.

In order to model the effects of different aspects of the structure of the descriptions on reading times, a multiple linear regression model was fitted to the data. The purpose of building a regression model is to arrive at a better articulated description of the functions that relate reading time to various psychological processes — to reveal how reading time is related to the amounts of different types of information being held in memory.

The independent variables used were as follows. All the variables from the Stenning, Shepherd and Levy model were used as candidate variables in developing the new model. All these 'load' variables take values related the to number of properties known of the current referent. MISLOAD was defined as the number of properties known of the current referent that mismatched with the other individual. MATLOAD was defined as the number of matched properties. NEUTLOAD was defined as the number of properties whose matching status is unknown. LOCALMIS is a binary variable with value one when a property mismatch is determined by the current sentence.

To investigate differential effects of the primary/secondary status of individuals, a number of variables were entered into the regression procedure along with versions of themselves distinguished by status. Thus, MATLOAD and MISLOAD were distinguished by whether the referenced individual was primary or secondary (PRMAT/SECMAT, PRMIS/SECMIS). NEUTLOAD only ever takes a positive value on a secondary individual at one sentence in one mode (Mode 5, Sentence 7): the secondary individual, by definition, generally lags behind the primary one. This one position does not allow reliable estimation of primary/secondary differences in NEUTLOAD coefficients.

To investigate the effects of referential change, three more variables were defined. CONTOUR is a binary variable with value one if the current sentence had switched reference from the previous sentence's referent, and otherwise with value zero. FOREGROUND is a variable with value equal to the number of properties previously known of the individual to which reference has just switched, or zero if reference continues. FOREGROUND was differentiated into two variables (PRFORE and SECFORE) by the primary/secondary status of the referenced individual.

Stenning, Shepherd and Levy, showed that the net effects of word length were limited to a few sentence positions in their predictable texts, and they made no attempt to incorporate word length variables into their regression model. Once primary/secondary differences and word-length effects related to change of reference had been discovered, a more systematic search for word-length effects was made by defining variables which took as their values the number of syllables accumulated in the descriptions of individuals at various points in reading texts.

Basically two variables were defined: REFSYL and NONREFSYL took as their values the number of syllables in the accumulated description of the referenced and non-referenced individuals respectively. So, for example, if a subject was reading the sentence *The dentist was Swiss* in a text in which the information so far had been that there was a dentist and a fat Polish bishop, the values of REFSYL and NONREFSYL would be 3 (Swiss/den/tist) and 5 (fat/Pol/ish/bish/op) respectively. These variables were further differentiated by whether the referent was primary or secondary, whether the syllables were of matched or mismatched properties, whether reference had just switched (i.e. CONTOUR = 1) and whether the syllables had accumulated since the start of the text or during the current run of reference. The syllabic variables took values between 0 and 11 (the maximum number of syllables in any individual's complete description in the texts used). The same distinctions were used to define the accumulation of properties on referenced and non-referenced individuals.

Table 3 shows two example paragraphs with the values of some of the independent load variables, assigned at each point. The examples are a text in Mode 4 with mismatched attributes on the first, third and fourth dimension, and one matched attribute on the second dimension, and a text in Mode 7 with mismatched attributes on the first, second and third dimension, and matched attributes on the fourth dimension.

(TABLE 3 ABOUT HERE)

In order to allow the data to determine the shape of the functions, dummy variables were used to represent each of the property load variables MISLOAD, MATLOAD, NEUTLOAD and FOREGROUND. So, for example, MISLOAD is represented by the four dummy variables MIS1, MIS2, MIS3 and MIS4. (See Draper and Smith, 1981 for a discussion of this use of dummy variables in multiple regression). Briefly, the N levels of a pseudo-continuous variable are represented by N-1 binary variables, each defined so they take the value 1 at their unique level of the parent variable, and otherwise the value 0. Since our empirical aim is to describe what shape these functions take, we did not wish to constrain them to be linear. We did not similarly define binary dummy variables for all the values of the syllabic variables. There were too many values, but more importantly, there are theoretical reasons for imposing a linear function on these variables (see Baddeley, 1986).

Table 4 presents the model of Experiment 1's reading time data, showing the coefficients of the significant variables chosen using stepwise multiple linear regression. The variance accounted for is comparable to the regression models in Stenning, Shepherd & Levy (1988) where it was shown that the pure error (Draper & Smith, 1981) in their data accounted for nearly 90% of the total variance.

The following convention for naming variables is used in Table 4: REF=referenced, NON-REF=non-referenced, PR=primary, SEC=secondary, SW=switched reference, RUN=accumulation since reference switch, SYL=syllables, PROP=properties. For example, NON-REFSECSWSYL refers to the number of syllables in the secondary individual's description when it is the non-referenced individual after a switch of reference.

(TABLE 4 ABOUT HERE)

Discussion of the Regression Model

The regression model developed for the present data is a close relative of that developed by Stenning, Shepherd and Levy for data from a much simpler experimental situation. At a first approximation, one can describe the present model as being derived from the old one by adding variables to take account of new processes arising from new demands posed by the more complex texts. Also, it appears that changes in strategy for the unpredictable texts have caused MATLOAD and LOCALMIS to drop out of the regression equation.

The Process of Changing Reference

What do the additions to the model tell us about the processes involved in referential change? Table 5 summarises the relevant variables' coefficients.

(TABLE 5 ABOUT HERE)

The regression model shows that switching reference unpredictably to a primary individual with more specification is slower than to one with less specification. There is a steadily increasing cost with number of properties for a switch to a primary individual but a small or negligible cost for switching to a secondary individual however many properties are known. Changing reference to a *new* individual actually speeds processing (relative to continuing with the established reference). Although this finding appears in conflict with the results of Kieras (Kieras & Just, 1984), it should be remembered that in our constrained texts, a great deal is known about the introduction of the second individual before it has occurred.

Mode 10 was designed to test whether the long delays on returning reference to a nearly complete primary individual would appear when the primary individual was the second one introduced. Long reading times appear here comparable to those associated with returning to primary individuals introduced first in Modes 2 and 4. This constitutes a further justification of the distinction between primary and secondary status: when status changes, the newly defined primary individual shows this most distinctive characteristic of primary individuals.

Word-length Effects and the Role of Articulatory Rehearsal

Table 6 summarizes the word length effects. Negative coefficients in the model represent accelerations in processing just as positive coefficients signify additional processing. Naturally, the model will never predict a negative reading time because the contribution of the negative coefficients is always far outweighed by positive effects. The largest word-length effect is the acceleration in reading time in proportion to the number of syllables in the secondary individual's description when there is a switch of reference to the primary individual. In general the pattern of effects when reading about the primary individual have remained the same as in the pilot experiment. When reference is continued to the primary individual, each syllable in both individuals' descriptions contributes a positive time and each syllable in the primary individual's description accumulated since the last switch of reference contributes a small but significant amount. When reference has switched to the primary individual the word length effects of the primary individual's description are the same as those when reference has continued. However, as described above, the secondary individual's description accelerates reading.

(TABLE 6 ABOUT HERE)

One of our goals is to provide an account of the interaction between short term memory and long term memory and in order to do this we must provide a clear interpretation of this negative effect of the length of the description of the secondary individual on the time spent reading a subsequent sentence about the primary individual. There are two candidate interpretations that we wish to distinguish. The first interpretation is the one, already described, which sees the negative coefficient as representing an acceleration of the cognitive machinery used to process the primary individual. The second interpretation sees the negative effect representing abbreviation of the processing of the primary individual rather than straightforward acceleration — perhaps rehearsing a reduced part of the primary individual's description. Of course, the negative effect could be due to a mixture of both processes. However, we could not distinguish between these two possibilities using regression modeling alone so we designed Experiment 2 in which we recorded subjects' overt rehearsal protocols while they performed an MIT (see Ericson & Simon, 1984, for a justification of this technique). The overt rehearsal would then allow us to observe subjects' rehearsal when there is a switch of reference to the primary individual. Experiment 2 was also designed to replicate the pattern of recall confusions for Modes 1 – 7 and test the predicted confusions for the three new modes (Modes 8 – 10) used in Experiment 1. In order to do this we employed the cued-recall procedure used in the pilot experiment.

Experiment 2

Method

Materials

The vocabulary used in the experiment appears in Appendix . The vocabulary set contains 48 words which were split into four groups of twelve, each denoting occupation, nationality, stature and temperament. The same ten modes that were used in Experiment 1 were used in this experiment and the paragraph construction was also the same.

Design

Each subject saw 40 texts which were split up into four sessions of ten texts which will be referred to as *blocks*. Mode was crossed with block, and matchtype was balanced by subject. The sequence of mode and matchtype was randomised. The full design consisted of the following factors (and levels): Mode (10 levels), Individual (2 levels), Property (4 levels) and Block (4 levels).

All factors were within-subject factors and were fully crossed. The reading time and any vocalizations (which were presumed to be overt rehearsals) for each sentence were recorded. At the end of each text the subject was cued to recall one or other of the individuals first as in the pilot experiment. The cue was balanced across blocks.

Procedure

Each subject was tested individually with the four blocks taking between 40 minutes to just over an hour to complete. Seventeen volunteer subjects took part in the experiment. The procedure was the same as the procedure in Experiment 1 with the addition of the instructions relating to overt rehearsal.

At the end of Block 1 subjects were requested to vocalize any rehearsals they were making for the next two blocks. For the final block subjects were given the option of either continuing to rehearse aloud or of reverting to their original strategy (if this was different from overt rehearsal). Subjects were given no instructions regarding particular rehearsal strategies.

The materials were presented on a BBC model B microcomputer which also recorded the reading times. A video recorder with microphone was used to record the computer's output to its monitor and the subjects' vocalizations. This was to facilitate the logging of the vocal data by synchronizing it with the materials.

Results

Any overt rehearsal for each sentence were recorded as was the recall for each text. Only two subjects out of the 17 subjects made no overt rehearsals in block four.

Recall

As in the pilot data, it was found that in some modes there are more confusions about the order of introduction than there are in other modes. The modes in which they observed this confusion were modes 3, 5 ,6 and 7. The patterns of confusion were the same as in the pilot data for the modes that the two experiments had in common. We predicted that there would be confusions for the new mode 10 and this proved to be correct. A more detailed analysis of the recall data can be found in Nelson (forthcoming).

Overt rehearsal protocols

Pre-processing and logging subjects' verbalizations

The subjects verbalizations were not restricted in any way. However, the most relevant aspects of their protocols (within this experiment) was their use of the properties, presented to them in the experimental texts. Therefore the aim of the logging process was to record any mention of an experimental property within a subject's verbalizations.

Effects of switching and status

An ANOVA was carried out with subjects as a random factor, status and switch of reference as fixed factors with the number of properties of the referenced individual missed out divided by the number of properties known of the referenced individual as the dependent variable. Only data for the last three blocks was used. Status had a main effect (means 0.11 and 0.08 for primary and secondary respectively), $F(1,16) = 4.794, p < 0.05$. There was an interaction between status and switch of reference, $F(1,16) = 16.556, p < 0.001$. See Table 7 for the means.

(TABLE 7 ABOUT HERE)

The table clearly shows that when there is a switch of reference to the primary individual more properties are missed out than when there is a continuation of reference. This confirms one of the possibilities admitted by the negative word-length effect found in Experiment 1, which was that the saving in time was acheived by not rehearsing some of the properties of the primary individual. Table 7 also shows that for the secondary individual more properties are missed out when there is a continuation of reference compared to a switch of reference. This result confirms the negative word-length effect for a continuation to the secondary individual (see Table 6) found in Experiment 1.

General Discussion

Simply introducing one circumscribed type of unpredictability into this task has radical effects on performance. An asymmetry between individuals appears which is chiefly evidenced by differences in the processes of referential change, but also by indications of competition for resources. Both the asymmetry and the competition for processing time between the two individuals only occurs in unpredictable texts.

The increase of time taken to shift reference to primary individuals with their increasing specification suggests that the whole of their representation (or at least some constant proportion of it) has to be reactivated. This is an observation of an analogue of Anderson's (1983) fan-effect in an on-line discourse processing task where representations are under construction rather than interrogation. However, the combination of these present results with the previous observation of the absence of switching effects, casts doubt on Anderson's explanation in terms of spreading activation. This model has difficulty both with the absence of effects when texts are predictable, and with the asymmetry of effects in the unpredictable case.

Why should unpredictable change of reference cost time when predictable change does not? Perhaps prediction allows processes of shifting attention to start earlier and to run in parallel. Perhaps when a shift of reference is expected, these processes can begin before processes of representing the current attribution to the old referent are finished. It is however unlikely that processes taking quite considerable and very different amounts of time for shifting attention to individuals of different complexity would all be neatly absorbed leaving no trace in the predictable task. Another alternative is that the predictable texts allow the subject to avoid altogether carrying out construction of some representations. This latter alternative is the explanation which we will explore here.

Looking at the patterns of binding in the predictable and unpredictable texts, it is striking that there is a simple relationship between superficial facts about the predictable texts and their patterns of binding, and that this simple relationship is disturbed in the unpredictable texts.

In Stenning, Shepherd and Levy's texts, for both their modes (I $\times$ I and P $\times$ P), any first mentioned value on a property dimension is a property of the first individual, and any second mentioned value is a property of the second individual. If subjects can remember that *fat* came before *thin* then they know that the first individual was fat and the second introduced individual was thin. Representations of low level temporal facts — in particular within-dimensions ordering facts — about the input suffice as representations of the semantic structure.

Our answer to the question, why does switching reference predictably cost no time, is that predictable texts map directly onto representations of binding whereas unpredictable ones do not.

Why should there be an asymmetry between primary and secondary individuals in unpredictable texts but not in predictable ones? The proposed theory of how the representation of binding is distributed over a set of facts about within-dimension ordering of properties suggests the outline of an answer. Distinguishing primary and secondary individuals re-establishes a mapping between within-dimension-order and semantics, though a slightly more complicated one than was available with the predictable texts. The original mapping from within-dimension order to first/second introduced individual is replaced by a mapping from within-dimension order to primary/secondary individual. Complicating the interpretation (from simple order to primary/secondary status) allows the same primitive information about within-dimension order to be used in more general circumstances, but the extra complexity makes itself apparent in asymmetries between the individuals.

Let us define 'leading property' on a dimension as the first occurring, and 'trailing property' as the second occurring. In the present experiments, there are many cases (about a quarter of all sentences) in which the leading property is attributed to the *second* introduced individual and the trailing property is attributed to individual introduced *first*. But there are only two cases in eighty in which the *secondary* individual 'gets in front of' the *primary* individual (Mode 5, Sentence 7 and 8). The distinction between primary and secondary individuals restores the transparency of the relation between superficial property ordering relations and underlying semantic relations.

This restoration of transparency can be achieved only at the expense of some reorganisation when a change of status occurs. Properties represented as leading or trailing prior to the change must be re-represented if 'leading' and 'trailing' are to be uniformly interpreted as applying to the redefined primary and secondary individuals. This is made more plausible by the fact that such changes of status generally occur early in the processing of these texts.

The original clue to the distinction of status was the prevalence of confusions of identity in answering questions about texts where status had changed. We noted that in modes in which primary status has shifted from one individual to another, cueing recall by order of introduction lead to confusions of identity. Confusion between *primary* and *first introduced*, and between *secondary* and *second introduced* individuals is what one would expect if primary/secondary status is encoded in terms of information about sequence within-dimension. The fact that there is no evidence for the primary/secondary distinction in Stenning, Shepherd and Levy's data where it is unnecessary to restore a transparency which has not been disturbed, supports this interpretation.

Our results clearly show an asymmetry between the two individuals in the unpredictable texts used in these experiments. This asymmetry is related to "semantic" properties as well as superficial word-length and there appears to be a tension between superficial and semantic processing. Earlier results reported in Stenning, Shepherd and Levy (1988) support a distributed theory of binding and show that predictable texts do not evoke these asymmetries. Stenning and Levy (1988) provide a PDP model of distributed binding and show that it accounts for earlier experimental data. In order to reconcile the differences between predictable and unpredictable texts we hypothesise that within-dimension order explains the imposition of asymmetry. Our experimental results support this hypothesis because our definitions of primary and secondary were determined by the data and those definitions correspond to the mapping between surface order and semantics. Further, the incorporation of sequence into our theory of distributed binding will simply mean adding within-dimension order information to the ensemble of low-level facts already postulated.

If this hypothesis is correct, any textual order which disrupts the simple mapping from surface order to semantics can be expected to lead to extra processing. Patel (1990), Experiment 2 explored the importance of congruence of order of dimensions across the two individuals. In the following schematic example text, although reference alternates regularly between individuals in property by property order, the dimensions of description for the two individuals diverge. The *stature* of the first introduced individual precedes the *stature* of the second, whereas the *temperament* of the second precedes that of the first.

> Chef. Vet. Chef short. Vet happy. Chef Welsh. Vet short. Chef sad. Vet Welsh.

This means that the sequence of properties within a dimension no longer maps onto the reference of the properties in a transparent way. The effects on reading time are complex, but are as disruptive as unpredictable reference. It appears that any disruption of the mapping of surface ordering onto semantics makes processing more difficult, however that disruption occurs.

The regression model (see Table 4 above) shows that the process of shifting reference to a primary individual takes roughly the same amount of time as the process of adding a new unresolved or mismatched property to the representation of an individual specified by the same number of properties. That is to say, to 'restore' an individual to attention takes about as much time as to add a new property to the representation of a referent already attended to. Whether the content of these two processes overlap or not, this is evidence that both access the whole of the known specification. Doubtless there are limits on this process when a large amount is known about an individual, but this is an important result for individuals of the complexity used here (and in most psycholinguistic experimentation). At least where memory is required, adding new information is not done by linking it to some identifier which remains of constant complexity, but by accessing the whole of what is known, or at least material in proportion to the amount known.

MIT texts are not naturalistic and the question therefore arises of how interpretations of these results transfer to conclusions about general discourse processing. Naturalistic texts contain both predictable and unpredictable sequences. We note that it would be unusual to depart from a predictable sequence (where one is available) without some reason (witness Cicero's advice on sequencing). But such sequences are often not available and our observations of the strategy of making a primary/secondary distinction exemplifies one coping strategy. The other main aspect of unnaturalness is the high memory load imposed by the arbitrary bindings in this task. First we observe that subjects perform better on this task, despite its high memory load, than on list recall tasks. We believe this is due to the high degree of semantic structure available in MIT texts which subjects can readily exploit. Secondly, naturalistic language (e.g. instruction manuals, regulations and technical literature) often poses much higher memory loads than typical narrative experimental texts. Thirdly, it is a methodological necessity to overload memory in order to study it but the results nevertheless can be indicative of the structures which underpin more naturalistic processing.

The two experiments described in this paper have revealed some important facts about the processing of representations of individuals. First, the time taken to read about an individual is related to the number of properties already known about that individual. Second, the effect of switching reference is also a function of the status of the individual to which reference is switched, and for switches with primary 'destination' individuals it is a function of the amount known about that individual. Lastly, there is an *acceleration* in reading when switching to a primary individual which is a positive function of the amount known of the non-referenced individual and this is the only position in which the amount known about non-referenced individuals affects reading time. We have offered explanations of these facts which are derived from a broad evidence base which gives them strong support. The regression model made predictions about rehearsal and patterns of confusions in cued recall which were confirmed in Experiment 2 which used a quite different measure.

The asymmetry between primary and secondary individuals that emerges in the present experiments is complementary to results from the anaphor resolution literature concerning the special status of a single thematic subject in narrative prose. It is commonplace that narrative tends to be organised around a single individual, at least within any episode. This privileged individual can be distinguished from lesser contemporaries by a variety of devices, such as being introduced first into an episode, and especially by being dubbed with a proper name (Sanford, Moar & Garrod, 1988). A character given a proper name can be accessed more quickly when referred to by pronouns than can the same character when she is described by the role she plays in the narrative. The individual thus privileged by whatever means becomes the focus of inferential elaboration (Garrod & Sanford, 1988).

Our results are interesting in this context precisely because no explicit devices are used to privilege an individual, yet asymmetries emerge between the processing of individuals. As long as one is dealing with rich naturalistic texts like those of Garrod and Sanford (1988), it is hard to tell whether differences between the representations of characters are a result of differences in our knowledge of them, or whether these differences are imposed by the computational characteristics of working memory. The present results suggest that the latter explanation should be taken seriously. If an explanation can be found for why working memory should privilege a primary over a secondary character even in our impoverished materials, this might go some way to explaining why narrative text is organised as it is.

The present theory that binding is represented by distributing facts about the temporal sequence of pairs of textual fragments potentially has wide ranging implications for cognitive processing. Binding is among the most basic components of any scheme of knowledge representation. Binding is not easy to achieve in distributed processing systems which cannot concatenate strings of symbol tokens and copy them from place to place. Language is a distinctive cognitive accomplishment because it consists of a system with *syntax* — a system in which positions in open-ended sequences have significance. Such high-level accomplishments rest on the ability to set up large-scale patterns of bindings between re-orderable elements on the basis of single exposures. The present theory proposes that these structures can be implemented on the basis of primitive representations of the sequence within pairs of elements, and these in turn, in a fully developed theory, would be represented on the basis of contentful associations. Explaining how binding is implemented on primitive mechanisms not readily adapted to it, is one of the central problems of cognitive science.

Whatever balance of phonemic and semantic processing is responsible for these effects, they are evidence of a tension between attended and unattended individuals, a tension which operates asymmetrically between primary and secondary individuals. When the primary individual is referenced, there is a tendency to accelerate in proportion to the amount known about the secondary individual. When the secondary individual is being processed, there is a tendency to spend time also processing the primary individual in proportion to how much is known of it.

These two observations cannot be explained simply in terms of the primary being more important than the secondary — while this might explain the intrusion of processing of the primary individual while the secondary is referenced it cannot explain a willingness to abbreviate processing of the primary when the secondary is referenced. There must be some qualitative difference in the nature of representations involved. Discovering the nature of this difference is the most pressing need for development of an account of the primary/secondary distinction and ultimately of a computational model of these processes.

Footnote

[1] There are not many single syllable nationalities in English (for morphological reasons). *Welsh* was used as a high frequency item, although its rating in Francis and Kucera (1982) is low. It is almost certainly a high frequency item for Scottish speakers as opposed to American writers. Also, *Czech* was used as a low frequency item although it had no rating in Francis and Kucera (1982) because *Czechoslovakia* was given a rating of 4 samples out 500.

References:

Anderson, J. R. (1983) *The Architecture of Cognition.* Cambridge, Mass.: Harvard University Press.

Baddeley, A. D., (1986) *Working Memory.* Oxford: Oxford University Press.

Campbell, R. and Butterworth, B. (1985) Phonological dyslexia and dysgraphia in a highly literate subject: a developmental case with associated deficits of phonemic processing and awareness.. *Quarterly Journal of Experimental Psychology,* **37A**, 435–76.

Draper, N. R. and Smith, H. (1981) *Applied Regression Analysis,* 2nd Edition. Chichester: John Wiley and Sons.

Ericson, K. A. and Simon, H. A. (1984) *Protocol analysis: verbal reports as data.* Cambridge, Mass: MIT Press.

Ehrlich, K. and Johnson-Laird, P. N. (1982) Spatial descriptions and referential continuity. *Journal of Verbal Learning and Verbal Behavior,* **21**, 296–306.

Francis, W. N. and Kucera, H. (1982) *Frequency Analysis of English Usage: Lexicon and Grammar.* Boston, Mass.: Houghton Mifflin.

Garrod, S. C. and Sanford, A. J. (1988) Thematic subjecthood and cognitive constraints on discourse structure. *Journal of Pragmatics,* **12**, 519–534.

Johnson-Laird, P. H. (1975) Models of deduction. In Falmagne, R. J. (ed.) *Reasoning: Representation and Process in Children and Adults.* Hillsdale, N.J.: Lawrence Erlbaum Associates.

Kieras, D. E. and Just, M. A. (eds.) (1984) *New Methods in Reading Comprehension Research.* Hillsdale, N.J.: Lawrence Erlbaum Associates.

Levy, J., Nelson, A. W. R. and Stenning, K. (in prep.) Performance of the Memory for Individuals Task under conditions of articulatory suppression.

McGonigle, B. and Chalmers, M. (1986) Representations and Strategies During Inference. Chapter 6 in Myers, T. F., Brown, E. K. and McGonigle, B. (eds.) *Reasoning and Discourse Processes.* London and San Diego: Academic Press.

Nelson, A. W. R. (forthcoming) An investigation of asymmetry in the process of referential change in the construction of representations of individuals. PhD Thesis, Centre for Cognitive Science, University of Edinburgh.

Patel, M. J. (1990) Human text processing and models of knowledge representation. PhD Thesis, Centre for Cognitive Science, University of Edinburgh.

Sanford, A. J., Moar, K. and Garrod, S. C. (1988) Proper names as controllers of discourse focus. *Language and Speech,* **31**, 43–56.

Stenning, K., Patel, M.J. and Levy, J. (1987) The 'Binding Problem' in human memory: some effects of referential discontinuity on the construction of representations for individuals.

Research Paper EUCCS/RP–18, Centre for Cognitive Science, Edinburgh, December 1987.

Stenning, K., Shepherd, M. and Levy, J. (1988) On the construction of representations for individuals from descriptions in text. *Language and Cognitive Processes*, **2**, 129– 164.

Stenning, K. and Levy, J. (1988) Knowledge-rich solutions to the binding problem: a simulation of some human computational mechanisms. *Knowledge Based Systems*, **1**, 143–152.

Vallar, G. and Baddeley, A. D. (1984) Fractionation of working memory: Neuropsychological evidence for a phonological short-term store. *Journal of Verbal Learning and Verbal Behavior*, **23**, 151–161.

Yates, F. A. (1966) *The art of memory.* London: Routledge and Kegan Paul.

Appendices

Vocabulary set for Experiment 1

(Table 8 about here)

Vocabulary set for Experiment 2

(Table 9 about here)

	Sentence Position							
Mode	1	2	3	4	5	6	7	8
1	a 1	a 2	a 3	a 4	b 1	b 2	b 3	b 4
2	a 1	a 2	a 3	b 1	a 4	b 2	b 3	b 4
3	a 1	a 2	b 1	b 2	b 3	b 4	a 3	a 4
4	a 1	b 1	a 2	a 3	b 2	a 4	b 3	b 4
5	a 1	b 1	b 2	a 2	a 3	b 3	a 4	b 4
6	a 1	b 1	b 2	b 3	a 2	a 3	b 4	a 4
7	a 1	b 1	b 2	b 3	b 4	a 2	a 3	a 4
8	a 1	a 2	b 1	a 3	a 4	b 2	b 3	b 4
9	a 1	a 2	b 1	b 2	a 3	a 4	b 3	b 4
10	a 1	b 1	b 2	b 3	a 2	b 4	a 3	a 4

Table 1: Sentence sequences in each mode (letter denotes individual and number denotes property).

Property	Individual 1				Individual 2				Mean
	1	2	3	4	1	2	3	4	
Mode 1	1.57	1.79	1.90	2.81	1.69	1.75	1.71	2.21	1.93
Mode 2	1.37	1.59	1.75	3.20	1.61	1.87	1.73	2.12	1.90
Mode 3	1.62	1.68	2.08	2.24	1.28	1.85	2.15	2.42	1.91
Mode 4	1.43	1.84	2.06	3.32	1.33	2.11	2.19	2.02	2.04
Mode 5	1.42	1.87	2.10	2.17	1.27	1.93	2.17	2.31	1.90
Mode 6	1.39	1.95	2.13	2.31	1.19	1.88	1.87	2.19	1.86
Mode 7	1.49	1.90	1.75	2.01	1.30	1.82	1.90	2.43	1.83
Mode 8	1.47	1.67	2.67	2.76	1.28	2.10	1.83	2.11	1.99
Mode 9	1.54	1.74	2.06	2.49	1.32	1.85	1.97	1.97	1.86
Mode 10	1.40	2.01	2.31	2.09	1.26	1.87	1.86	3.02	1.97
Mean	1.47	1.80	2.08	2.54	1.35	1.90	1.94	2.28	1.91

Table 2: Experiment 1 mean reading times (sec) as a function of individual, property and text mode

Text	MIS-LOAD	MAT-LOAD	NEUT-LOAD	CONTOUR	FORE-GROUND
			Mode 4 +−−		
There is a dentist	0	0	1	0	0
There is a bishop	1	0	0	1	0
The dentist is Swiss	1	0	1	1	1
The dentist is thin	1	0	2	0	0
The bishop is Swiss	2	0	0	1	1
The dentist is tall	2	0	2	1	3
The bishop is fat	2	1	0	1	2
The bishop is short	3	1	0	0	0
			Mode 7 −−+		
There is a dentist	0	0	1	0	0
There is an bishop	1	0	0	1	0
The bishop is Polish	1	0	1	0	0
The bishop is fat	1	0	2	0	0
The bishop is short	1	0	3	0	0
The dentist is Swiss	1	1	0	1	1
The dentist is thin	2	1	0	0	0
The dentist is short	3	1	0	0	0

Table 3: Two examples of values taken by some of the variables in the regression model.

Variable	Coeff(sec)	St. Err.	F ratio
INTERCEPT	1.58		
FREQUENCY	0.09	0.02	19.84
NEUT2	0.21	0.04	32.90
NEUT3	0.47	0.06	67.45
NEUT4	1.32	0.11	151.67
MIS1	0.26	0.05	29.34
MIS2	0.60	0.06	117.47
MIS3	0.76	0.07	133.99
MIS4	0.69	0.10	44.95
CONTOUR	-0.58	0.07	74.38
FORE1	0.15	0.04	10.98
PRFORE1	0.83	0.12	46.13
PRFORE2	1.47	0.11	165.34
PRFORE3	1.90	0.13	214.67
NON-REFPRSWPROP	0.17	0.02	65.43
NON-REFSECSWPROP	-0.21	0.06	11.32
REFSYL	0.05	0.01	23.97
NON-REFSYL	0.07	0.02	9.83
NON-REFPRSYL	-0.09	0.02	27.83
NON-REFSECSWSYL	-0.15	0.04	16.61
REFRUNSYL	-0.04	0.01	12.93

Table 4: Summary of reading time regression model for Experiment 1 ($R^2 = 0.121, F(20, 12779) = 87.99$).

Status of Individual	Continues Reference	No. Properties of New Referent			
		0	1	2	3
Primary	0	-0.58	0.99	1.47	1.90
Secondary	0	-0.58	0.15	0	0

Table 5: Summary of effects of referential change in the regression model: total coefficients (seconds) for each property by continuity of reference and number of properties previously known of new referent and by primary/secondary individual

	Reading about Primary		Reading about Secondary	
	Referenced	Non-referenced	Referenced	Non-referenced
Reference continued	0.05/0.01*	0.06	0.05/0.01*	-0.03
Reference changed	0.05/0.01*	-0.09	0.05/0.01*	-0.03

Table 6: Summary of syllabic effects in the regression model: coefficients (seconds) for each syllable of description accumulated since the beginning of the text, on referenced and non-referenced individuals, by continuity of reference and by primary/secondary individual Note: In cells marked * there is a difference in the coefficients for syllables accumulating before the current reference was established (shown on the left), and those accumulating since (shown on the right)

Currently referenced	Reference	
Individual	Continued	Switched
Primary	0.05	0.11
Secondary	0.18	0.06

Table 7: Number of properties missed out divided by number of properties known by status and switch of reference for Blocks 2 to 4

Dimension 1 Profession	Dimension 2 Nationality
secretary (74) / doctor (108)	Russian (34) / English(47)
judge (41) / maid (21)	Welsh (4) / Greek (20)
teacher (52) / clerk (22)	American (197) / French (58)
singer (12) / miner (7)	Canadian (6) / Polish (7)
nun (7) / vet (1)	Dutch (6) / Czech (-)
actress (7) / chef (7)	Portuguese(2) / Swiss (6)

Dimension 3 Personality	Dimension 4 Physical characteristics
generous (23) / modest (26)	gentle (24) / heavy (86)
bright (59) / sad (25)	young (187) / old (257)
alert (24) / dull (26)	healthy (21) / weak (23)
gloomy (3) / cheerful (10)	graceful (9) / awkward(8)
shy (11) / gruff (3)	slim (9) / stout (2)
eccentric (8) / sane (8)	hairy (5) / bald (4)

Table 8: Vocabulary set for Experiment 1 by dimension. Numbers in parentheses are frequency of occurrence out of the 500 sample from Francis & Kucera's (1982) word count.

Dimension 1 Profession	Dimension 2 Nationality	Dimension 3 Personality	Dimension 4 Physical characteristics
dentist/baker	Chinese/Polish	bright/dumb	cheerful/sad
dentist/chef	Chinese/Swiss	clever/dim	daring/meek
doctor/priest	Dutch/Czech	clever/stupid	daring/timid
doctor/vicar	Dutch/Russian	clever/thick	friendly/hostile
judge/bishop	French/Greek	fat/skinny	happy/gloomy
judge/monk	French/Spanish	fat/thin	happy/sad
nurse/priest	German/Greek	greedy/clumsy	nice/nasty
nurse/vicar	German/Spanish	hungry/thirsty	normal/mad
teacher/bishop	Swedish/Czech	rich/poor	sane/insane
teacher/monk	Swedish/Russian	smart/scruffy	sane/mad
vet/baker	Welsh/Polish	wealthy/poor	strong/weak
vet/chef	Welsh/Swiss	young/old	tall/short

Table 9: Vocabulary set for Experiment 2 by dimension