THE SYNTACTIC DIFFICULTY OF JAPANESE SENTENCES

So Hiranuma

Thesis submitted for the degree of Doctor of Philosophy

UNIVERSITY COLLEGE LONDON

2001
Abstract

This thesis investigates the various syntactic sources of difficulty or ease in processing Japanese sentences. The investigation utilises the theory of Word Grammar (Hudson 1984, 1990), within which dependency distance has been developed in order to measure syntactic difficulty. My main contention is that dependency distance is just one source of difficulty, so I argue against the common idea that syntactic difficulty equals syntactic complexity and that difficulty has a single syntactic explanation.

After presenting the nature of syntactic difficulty and complexity and techniques of measuring syntactic difficulty in the first chapter, in the second chapter recent theories accounting for syntactic difficulty are reviewed. The third chapter outlines a model of a parser based on Word Grammar along with dependency distance. In the fourth chapter the status of 'words' in Japanese is discussed within the framework of Word Grammar, which seeks to express syntactic knowledge in terms of direct relationships between words, so as to resolve the special problem of function words that Japanese raises. The following two chapters provide preliminary experimental data. In the fifth chapter the first experiment tentatively shows serial order effect, word-length effect, chunking effect and the contribution of function words to working memory load. The experiments in the sixth chapter tentatively show that dependency distance has a positive relationship with memorability on immediate recall tasks and "chaining" and "overload" effects. The data in the seventh chapter gives experimental evidence tentatively showing the appropriateness of measuring dependency distance for Japanese in terms of morpheme as well as "head" effects and effects of word order. Spoken English and Japanese texts are analysed in the eighth chapter to make a comparison of the syntactic difficulty amongst the texts, in which the results tentatively show that in terms of dependency distance Japanese is no more difficult for syntactic processing than English.
Acknowledgements

I would sincerely like to thank my thesis supervisor Professor Richard Hudson. It goes without saying that this work would never have been started, continued and most importantly finished, were it not for him. It was Dick who first taught me linguistics as well as his own theory, Word Grammar, and introduced me to the field of language processing. Since then he has been an inspiring teacher who has guided my progress for the last several years with exceptional insights, encouragement and superb management. I will always remain grateful for the patience and humanity he showed particularly when I was unable to work for many months due to ill health.

I am also deeply indebted to Teruyuki Zushi, the principal at the Institute of International Education in London, who gave me willing permission to conduct the experiments at the institute. I would also like to mention my gratitude to Luisa Element, Joanna Conn, and Izumi Nakamura who were responsible for producing the English and Japanese texts which were analysed in the eighth chapter. My gathering of the textual and experimental data would have been impossible without the countless participants in the experiments and the recording sessions of the conversations. I gratefully acknowledge their commitments. My special thanks go to Dr Andrew Faulkner whose thoughtful advice on statistics was invaluable. I have also benefited greatly from the technical support of William Robinson and Toshihiko Kubota in keeping my computers in working order.

Finally and most importantly I would like to thank my parents for putting up with me and for not asking too often when this work was to be completed. I am particularly grateful for my father's financial support throughout my studies.
Contents

Title 1
Abstract 2
Acknowledgements 3
Contents 4
List of Tables 7
List of Graphs 9

1 Introduction 10
1.1 Syntactic difficulty and complexity 10
1.2 Dependency distance 11
1.3 Methods for measuring syntactic difficulty 13
1.3.1 Frazier and Rayner (1988) vs. Hudson (1998) 14
1.3.1.1 Sentential subject vs. extraposed sentential subject 18
1.3.1.2 Non-extraposed relative clause vs. extraposed relative clause 19
1.3.1.3 Subordinate clause-initial vs. main clause-initial 20
1.4 Some research questions 20

2 Recent theories of syntactic difficulty 22
2.1 Frazier and Rayner (1988) 22
2.2 Gibson (1991, 1995) 26
2.3 Hawkins (1994) 35
2.4 Gibson (1998) 44
2.4.1 Discussion 57
2.5 A summary of the views of short-term memory 59

3 A Word Grammar parser 65
3.1 Word Grammar 65
3.1.1 Default Inheritance 66
3.1.2 Overriding 68
3.1.3 Temporary words and permanent words 72
3.1.4 The Best Fit Principle 73
3.1.5 Word Grammar universal principles 78
3.2 Word-Grammar based parsing 87
3.3 Dependency Distance and Density 102
### List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1.1</td>
<td>Results of reading times for the experiments by Frazier and Rayner (1988)</td>
<td>16</td>
</tr>
<tr>
<td>Table 1.2</td>
<td>Results of the success rate for the experiments by Hudson (1998)</td>
<td>18</td>
</tr>
<tr>
<td>Table 1.3</td>
<td>Reading time per character and gaze durations for (1.1)</td>
<td>19</td>
</tr>
<tr>
<td>Table 1.4</td>
<td>Reading time and gaze durations for the relative clause sentences</td>
<td>20</td>
</tr>
<tr>
<td>Table 2.1</td>
<td>IC-to-non-IC and IC-to-word ratios of (2.16) and (2.17)</td>
<td>42</td>
</tr>
<tr>
<td>Table 3.1</td>
<td>Mean-distances and densities for (3.27), (3.28) and (3.29)</td>
<td>108</td>
</tr>
<tr>
<td>Table 4.1</td>
<td>Inflections of a full verb ‘tobu’ (to fly) and an auxiliary verb ‘masu’ (polite)</td>
<td>126</td>
</tr>
<tr>
<td>Table 4.2</td>
<td>Inflections of ‘full verb’, ‘auxiliary verb’ and ‘adjective’</td>
<td>137</td>
</tr>
<tr>
<td>Table 4.3</td>
<td>Notation for the Word Grammar word-classes, grammatical functions, and inflections for Japanese</td>
<td>138</td>
</tr>
<tr>
<td>Table 5.1</td>
<td>Characteristics of the sequences used in Experiment 1</td>
<td>143</td>
</tr>
<tr>
<td>Table 5.2</td>
<td>The number of Failures and Successes for each sequence</td>
<td>149</td>
</tr>
<tr>
<td>Table 5.3</td>
<td>Number of failures for each unit in all sequences used in Experiment 1</td>
<td>150</td>
</tr>
<tr>
<td>Table 5.4</td>
<td>The number of Failures and Successes for Category I</td>
<td>151</td>
</tr>
<tr>
<td>Table 5.5</td>
<td>The number of Failures and Successes for Category II</td>
<td>151</td>
</tr>
<tr>
<td>Table 5.6</td>
<td>The number of Failures and Successes for Category III</td>
<td>152</td>
</tr>
<tr>
<td>Table 5.7</td>
<td>The data for ‘shrt dic 1+2’, ‘auxdries 1+2’ in Category II and ‘long dic 1+2’ and ‘shrt d+p 1+2’ in Category III</td>
<td>153</td>
</tr>
<tr>
<td>Table 5.8</td>
<td>Protected t between ‘shrt dic 1+2’ and ‘auxdries 1+2’</td>
<td>153</td>
</tr>
<tr>
<td>Table 5.9</td>
<td>Protected t between ‘long dic 1+2’ and ‘shrt d+p 1+2’</td>
<td>153</td>
</tr>
<tr>
<td>Table 6.1</td>
<td>Characteristics for all the sentences in Experiments 2 and 3</td>
<td>162</td>
</tr>
<tr>
<td>Table 6.2</td>
<td>Dependency distance for all the sentences in Experiments 2 and 3</td>
<td>162</td>
</tr>
<tr>
<td>Table 6.3</td>
<td>Order of the presentation of the sentences</td>
<td>163</td>
</tr>
<tr>
<td>Table 6.4</td>
<td>Results for Subject Group A</td>
<td>165</td>
</tr>
<tr>
<td>Table 6.5</td>
<td>Results for Subject Group B</td>
<td>165</td>
</tr>
<tr>
<td>Table 6.6</td>
<td>Results for Subject Group C</td>
<td>165</td>
</tr>
<tr>
<td>Table 6.7</td>
<td>Results for the sentences (1) and (2)</td>
<td>169</td>
</tr>
<tr>
<td>Table 6.8</td>
<td>Number of failures for each word in (1) and (2)</td>
<td>170</td>
</tr>
<tr>
<td>Table 6.9</td>
<td>Results for the sentences (3) and (4)</td>
<td>181</td>
</tr>
</tbody>
</table>
Table 7.1  Dependency distance for all the stimuli in Experiment 4 187
Table 7.2  Set 1 and Set 2 utilised in Experiment 4 188
Table 7.3  Number of failures for each morpheme in all the sentences 189
Table 7.4  Total number of failures made by each subject 190
Table 7.5  Total number of failures scored for each sentence 191
Table 7.6  Total number of failures made by Group 1 (S1-S9) 192
Table 7.7  Total number of failures made by Group 2 (S10-S18) 193
Table 7.8  Percentages of failures and the order of heads 195
Table 8.1  Characteristics of the five texts 213
Table 8.2  Dependency distances of the five texts 214
Table 8.3  Results for words with multiple dependents 215
Table 8.4  Data for words with one or multiple dependents 215
Table 8.5  Data for words with one or multiple dependents in JT1, 2 and 3 216
Table 9.1  Effects on memorability found in the experiments 246
List of Graphs

Graph 5.1  Average value of failures for each unit in all sequences 150
Graph 6.1  Subject group A's performance on the sentence (1a) 171
Graph 6.2  Subject group A's performance on the sentence (1b) 172
Graph 6.3  Subject group A's performance on the sentence (2a) 172
Graph 6.4  Subject group A's performance on the sentence (2b) 173
Graph 7.1  Number of failures for each morpheme in all the sentences 190
1 Introduction

In the literature of theories of sentence processing, great attention has been paid to the relationship between syntactic decoding in comprehension and the computational resources of working memory utilised for processing. In particular, it has been taken for granted that syntactic complexity correlates with difficulty specifically caused by syntactic processing.

However, in section 1.1, specifically I will draw attention to the difference between syntactic complexity and difficulty. While Section 1.2 outlines dependency distance as a measure of potential syntactic difficulty, Section 1.3 will introduce some representative methods for quantifying syntactic difficulty, immediate recall and eye movement techniques in particular.

1.1 Syntactic difficulty and complexity

The goal of reading or listening is to derive meaning, a process which involves such different stages as identifying input words, accessing their meanings, assigning syntactic structures, building semantic representations, and constructing a representation of the whole text or conversation. In the study of sentence processing, various attempts have been made to account for why some syntactic constructions are harder to process than others, which is often referred to as 'processing load'. Some well-known hypotheses attribute processing overload to the number of incomplete parsed phrase-structure rules (Yngve 1960; Chomsky & Miller 1963; Miller & Chomsky 1963; Miller & Isard 1964; Abney & Johnson 1991), or locally unsatisfied X-bar (Lewis 1993) or Case-assignment (Stabler 1994) or thematic (Gibson 1991) relationships.

Within the theories of processing overload, the term 'processing difficulty' has been used as a virtual synonym of 'complexity' in terms of assessing the processing load of a sentence, but they nonetheless should be differentiated from each other. One good reason for drawing this distinction is that these terms should refer to different entities: Syntactic complexity should be relevant to such objective structural facts about sentences as syntactic structures and semantic or textual representations, and how complex they are, while on the other hand, processing difficulty must be concerned with the psychological strain of processing a sentence. The
distinction between complexity and difficulty is fundamental, so I shall use the term 'syntactic difficulty' strictly to refer to the processing difficulty of a sentence or a phrase caused in particular by its syntax. In contrast, 'syntactic complexity' refers only to aspects of the syntactic structure itself, although complexity itself is hard to define since there is no obvious way of measuring degrees of complexity.

Another reason for distinguishing the notions of complexity and difficulty is that they may conflict. For example, on the one hand an extraposed sentence such as (1.1b) (from Frazier & Rayner 1988) clearly has a more complex structure than its unextraposed counterpart (1.1a) because of the extra pronoun and the relation between this and the displaced clause. On the other, extraposition has the effect of easing processing difficulty, so there is a mismatch between syntactic complexity and difficulty.

(1.1a) That the both of the Siamese twins survived the operation is remarkable.

(1.1b) It is remarkable that the both of the Siamese twins survived the operation.

Earlier work assumed syntactic difficulty and complexity to be identical, to the extent of using the two terms interchangeably. For instance, it was claimed that some kinds of memory cost can be associated directly with nonterminal nodes (Frazier 1985; Frazier & Rayner 1988), as will be explained in section 2.1. Difficulty and complexity however, have been distinguished in more recent work, especially by Gibson (1998), for whom difficulty is caused by integrating an input into the current structure and retaining unintegrated structures in working memory (see section 2.4). My own work will also distinguish them.

1.2 Dependency distance

If syntactic complexity, as such, is not responsible for processing difficulty, what aspect or aspects of syntactic structure really are responsible? This is the central question addressed in this thesis. We cannot assume in advance that there is a single source of difficulty, though other theories tend to make this assumption. Instead the research described in later chapters shows that difficulty is directly related to at least two distinct aspects of
syntactic structure, which contribute in different ways to difficulty.

One of these is "dependency distance", which is easily defined in Word Grammar (Hudson 1984, 1990), a general theory of language, which finds its modern roots in the work of Tesnière (1959). Word Grammar recognises words as the only units of syntax, so the structure of a sentence consists entirely of dependencies between pairs of individual words which are connected by syntactic rules. The theory provides a simple basis for predicting syntactic difficulty: Dependency distance is the number of words between a word and the word on which it depends. The difficulty for a sentence is therefore predicted from the mean dependency distance - the mean distance value for all words in a sentence. (See the following section for examples.) The main idea of dependency distance is that linking each word as a dependent to some other word requires computational resources which are proportional to the distance between the words being linked since unintegrated words must be held active in working memory.

Dependency distance is only concerned with 'surface' dependency structure where one word depends on precisely one other word, because the purpose of syntactic processing is to complete the linking of words by dependencies in the surface structure. This is how dependency distance distinguishes syntactic difficulty from complexity, which has to take into consideration total dependency structure - 'surface' plus 'extra' structure. When one word depends on more than one other word (e.g. structure-sharing, extraction and extraposition), except for only one dependency compatible with the geometry of surface structure, all the other dependencies are relegated to the 'extra' structure. (See Chapter 3 for a more detailed explanation.)

As a measure of syntactic difficulty, dependency distance differs from other measures based on phrase-structure grammar; the processing load for a word-word relation is directly calculated by the length of the dependency in Word Grammar whereas within phrase-structure theories, syntactic difficulty is expressed in terms of the number of syntactic nodes. The lack of phrase nodes in Word Grammar gives an advantage to dependency distance over other node-counting measures. Take two three-word sentences (a) SVO and (b) SOV for instance, where S, V and O stand for subject, verb and object, respectively. (See diagrams (1.6) and (1.7) for the comparable dependency analyses for (a) and (b).) While the mean distance for (a) is
zero because S and O are both immediately next to V on which they depend, that for (b) is 1/3 owing to the subject dependency which is mediated by one word, namely O, resulting in the distance of 1. Hence, dependency distance predicts that (b) is syntactically more difficult to process than (a). In contrast, since the total number of phrase nodes for (a) and (b) should be the same, a simple node-counting measure cannot predict the difference in syntactic difficulty for (a) and (b).

The second aspect of syntactic structure which turns out to affect processing difficulty is the order of dependents. Dependencies are classified in Word Grammar along familiar lines as subjects, objects, topics, adjuncts and so on. Some orders of these elements are easier to process than others regardless of dependency distance. For example, in Japanese it is easier to process a structure whose topic is its first element than one in which some other element precedes the topic.

Chapter 3 will provide a theoretical framework in which these two aspects of syntactic structure are shown to be relevant to working memory.

1.3 Methods for measuring syntactic difficulty

There have been recent advances in methodologies for obtaining more accurate empirical data of human behaviour during on-line comprehension of language. Experimental investigation of sentence processing however, has concentrated predominantly on English in connection with theories of ambiguity resolution (MacDonald, Pearlmutter & Seidenberg 1994; Trueswell, Tanenhaus & Garnsey 1994; Mitchell 1994) or those of processing overload (Gibson 1991, 1998).

Two main experimental techniques, which are used to measure how much time is spent by readers in analysing different parts of sentences, have been popular, though others have also been used, such as on-line grammaticality judgement (Kurtzman, Crawford & Nychis-Florence 1991), speeded grammaticality judgement (Warner & Glass 1987; Ferreira & Henderson 1991), and "stop making sense" (Boland, Tanenhaus, Garnsey & Carlson 1995).

In self-paced reading (Just, Carpenter & Wooley 1982; Kennedy & Murray
1984; King & Just 1991), subjects read a computer display and press a key every time they are ready for a new word. Eye-movement tracking (Rayner & Pollatsek 1989; Traxler & Pickering 1996; Pickering 1996) tells us not only how much time subjects are fixing their eyes on different parts of a sentence, but whether or not and where they make regressive eye movements.

The experimental technique used in this thesis is immediate recall, which is a common technique to test short-term memory retention with a history in research of this type - the elicited imitation tasks of Miller & Isard (1964), Lado (1965), Epstein (1967), Cook (1979) and Bley-Vroman & Chaudron (1994). Wingfield and Butterworth (1984) also used a similar method in their experiments, where subjects were given control of a pause switch on the tape recorder and allowed to stop the tape-recorded passages whenever they wished to begin their reports. The immediate recall technique used for Hudson's short-term memory retention tests (Hudson 1998) will be dealt with below in comparison to the eye tracking technique used by Frazier and Rayner (1988) in their experiments.


In their paper, Frazier and Rayner (1988) reported two experiments utilising the eye movement technique to study how people look at and comprehend sentences when they read them by way of a complete record of the sequence of eye movements obtained for a sentence and the location and duration of the eye fixations sampled every millisecond. Twenty members of the University of Massachusetts community participated in Experiment 1 and twelve different members took part in Experiment 2. All of the subjects had normal uncorrected vision and they were naïve with respect to the purpose of the experiment. To ensure that subjects read for comprehension, they were periodically asked by the experimenter to report the sentence that they had just read. Their materials were minimal pairs of sentences with sentential subjects and sentences with extraposed sentential subjects (1.1), sentences with extraposed relative clauses and their non-extraposed counterparts (1.2), and subordinate-initial sentences and main-initial sentences (1.3). All the above three types of sentences were used in the first experiment, whereas in the second, only the pairs of the type (1.1) were used with priming, whereby sentences were presented immediately after a relevant context. Exemplified minimal pairs of the three types are
given with their dependency structures and distances. The primary results in terms of reading times were as shown in Table 1.1 below.

(1.1a)

That both of the Siamese twins survived the operation is remarkable.

mean distance = 1.636

(1.1b)

It is remarkable that both of the Siamese twins survived the operation.

mean distance = 0.916

(1.2a)

Any girl who takes karate lessons could break the table easily.

mean distance = 0.727

(1.2b)

Any girl could break the table easily who takes karate lessons.

mean distance = 0.727
Mary was laughing because the girls tickled the cat.

mean distance = 0.444

Because the girls tickled the cat, Mary was laughing.

mean distance = 0.888

KEY: s=subject; r=sharer; c=complement; o=object; a<=pre-adjunct; >a=post-adjunct; >x=extraposee

Table 1.1 Results of reading times for the experiments by Frazier and Rayner (1988)

<table>
<thead>
<tr>
<th>Exp 1</th>
<th>Type of sentence</th>
<th>Reading time</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1.1)</td>
<td>sentential subject</td>
<td>41.6 msec per character</td>
</tr>
<tr>
<td></td>
<td>extraposed sentential subject</td>
<td>34.5 msec</td>
</tr>
<tr>
<td>(1.2)</td>
<td>extraposed relative clause</td>
<td>43.7 msec</td>
</tr>
<tr>
<td></td>
<td>non-extraposed relative clause</td>
<td>37.3 msec</td>
</tr>
<tr>
<td>(1.3)</td>
<td>subordinate clause-initial</td>
<td>35.4 msec</td>
</tr>
<tr>
<td></td>
<td>main clause-initial</td>
<td>38.6 msec</td>
</tr>
<tr>
<td>Exp 2 (with appropriate contexts)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1.1)</td>
<td>sentential subject</td>
<td>34.3 msec</td>
</tr>
<tr>
<td></td>
<td>extraposed sentential subject</td>
<td>26.6 msec</td>
</tr>
</tbody>
</table>

We shall now consider the results in Table 1.1, comparing them to the report on the experiments conducted by Hudson (1998). His subjects were 20 students at the Phonetics and Linguistics department of University College London, all of whom were native English speakers. The experiments were held at various times between 1997 and 1998. The technique used in the experiments was immediate recall. The subjects were instructed to transcribe
a sentence immediately after it was dictated. Each sentence was read out once. Only the complete reproduction of a sentence was counted as success, and anything else as failure. Amongst others, the experimental sentences comparable to the ones tested by Frazier and Rayner fell into the types (1.1) and (1.2), although they do not constitute minimal pairs, for the paired sentences were not composed of exactly the same words. The dependency structures and distances of the relevant pairs (1.4) and (1.5) are illustrated below with the subjects’ performance measured in terms of the success rate for the sentences in Table 1.2.

(1.4a)

That every history student who failed in physics had passed in maths seems odd.

mean distance = 1.785

(1.4b)

It seems odd that every physics student who passed in maths had failed in history.

mean distance = 1.000

(1.5a)

Every linguist who decided to solve the data problem without a calculator did well.

mean distance = 1.071
Every student did well who tried to solve the maths problem without a computer.

mean distance = 0.428

<table>
<thead>
<tr>
<th>#</th>
<th>Type of sentence</th>
<th>W</th>
<th>D</th>
<th>% F</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1.4a)</td>
<td>sentential subject</td>
<td>14</td>
<td>1.785</td>
<td>70%</td>
</tr>
<tr>
<td>(1.4b)</td>
<td>extraposed sentential subject</td>
<td>15</td>
<td>1.000</td>
<td>15%</td>
</tr>
<tr>
<td>(1.5a)</td>
<td>non-extraposed relative clause</td>
<td>14</td>
<td>1.071</td>
<td>30%</td>
</tr>
<tr>
<td>(1.5b)</td>
<td>extraposed relative clause</td>
<td>14</td>
<td>0.428</td>
<td>10%</td>
</tr>
</tbody>
</table>

KEY: W=number of words; D=mean dependency distance; % F=percentage of failure

Before commenting on the results, a few points concerning the experimental techniques should be noted; eye movements and immediate recall are similar in that both are designed to explore a human language processing system; eye movements technique is more suited to determine precisely in which region of a sentence processing difficulties especially caused by local ambiguities occur by measuring average gaze durations associated with different regions in the sentence; the immediate recall technique, in contrast, aims at looking at how working memory requirements affect overall processing difficulties. Also eye-gaze is relevant to reading whereas immediate recall is suitable for hearing.

1.3.1.1 Sentential subject vs. extraposed sentential subject

As evidenced by both the results of Frazier and Rayner and of Hudson, the difficulty (as measured by reading times or recall errors) was greater for the sentential subject sentences than for the extraposed sentential subject sentences. As it stands, dependency distance correctly predicts the syntactic difficulty caused by the sentential subject clauses in these sentences.
Table 1.3 Reading time per character and gaze durations for (1.1)

<table>
<thead>
<tr>
<th>Region</th>
<th>Matrix</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>Matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1.1a)</td>
<td></td>
<td>That... twins</td>
<td>survived</td>
<td>the operation</td>
<td>is remarkable.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>42 (264)</td>
<td>52 (255)</td>
<td>28 (221)</td>
<td>53 (407)</td>
</tr>
<tr>
<td>(1.1b)</td>
<td></td>
<td>It is remarkable</td>
<td>that...twins</td>
<td>survived</td>
<td>the operation.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>24 (213)</td>
<td>29 (225)</td>
<td>33 (246)</td>
<td>47 (296)</td>
</tr>
</tbody>
</table>

Data are reading time per character (msec). Values in parentheses are gaze durations (msec).

Table 1.3 above based on Frazier and Rayner (1988: 254) shows that the reading time per character and gaze durations of the matrix region (is remarkable) of (1.1a) are the longest (Frazier and Rayner 1988: 253). We can assume that linking up the long dependency distance subject That with the matrix verb is caused considerable difficulty. I shall report in Chapter 6 a similar finding that an embedded relative clause region produced worse performance results in working memory retention tests with Japanese sentences.

1.3.1.2 Non-extraposed relative clause vs. extraposed relative clause

Contrary to the predictions of the branching hypothesis that a left-branching structure is more difficult to understand than its right-branching counterpart, Frazier and Rayner reported that the sentences with extraposed relative clauses took longer to read than their non-extraposed versions. This is less surprising from the Word Grammar viewpoint on the grounds that processing difficulty is principally assessed by distances of dependencies rather than their directions. So far as the exemplified sentences (1.2a) and (1.2b) are concerned, their mean dependency distances are equal (0.727), which suggests that the syntactic difficulty of (1.2a) ought to be theoretically equivalent to that of (1.2b). In turn, it could mean that the extraposition in (1.2b) does not serve its purpose of reducing processing difficulty, but rather it might have induced extra effort for the subjects to find the antecedent of the relative pronoun who since the sentence could be complete before this word. This suggestion seems to be borne out by the regions analysis for the relative clause sentences in which the relative clause region in the relative clause-extraposed sentences reveals the longest reading time and gaze durations, as can be seen in Table 1.4.
Conversely, in Hudson's data, the non-extraposed relative clause sentence (1.5a) and the extraposed counterpart (1.5b), demonstrate a strong correlation between their mean dependency distances and the subjects' performance; the success rate of (1.5a) is poorer than that of (1.5b). This is as predicted given the extreme difference between the mean distances: 1.071 for (1.5a) compared with 0.428 for (1.5b).

1.3.1.3 Subordinate clause-initial vs. main clause-initial

According to Frazier and Rayner's results for the pairs of type (1.3), main-initial sentences were slightly more difficult to process than subordinate-initial sentences. This result is hard to explain in terms of structural geometry, whether phrase structure or dependency. Whatever the explanation may be, this result shows that this aspect of syntactic structure is only one influence on difficulty, and may be outweighed by other influences such as the nature of the dependencies concerned and the order of the dependents.

One of the findings through the experiments by Frazier and Rayner was that left-branching per se did not create substantial syntactic difficulty, but rather long-distance dependency would critically increase the difficulty of syntactic processing. This finding seems to be supported by the immediate recall test results presented by Hudson, which suggest that the longer a dependency distance is, the more memory load it costs for its activation on working memory.

1.4 Some research questions

This thesis focuses on the effects of dependency distance and the order of dependents on processing difficulty. However before we can investigate dependency distance in Japanese we have to decide exactly how to measure it.
Modern theories of natural language processing regard words as the basis of comprehension processes (Marslen-Wilson & Tyler, 1980; Rayner & Pollatsek, 1989). Whilst words are mostly transparent entities for English, they are rather opaque for Japanese owing to its agglutinative nature. Bearing in mind that dependency distance is also a word-based measure, more specifically we need to consider what the unit is in Japanese which is equivalent to the word to which dependency distance can be applied. This question will be considered in Chapter 4.

Chapters 5 to 7 present experimental evidence on the relative difficulty of Japanese sentences with different syntactic structures which will allow us to separate the effects of dependency distance and word order.

Finally, given that the same measure, dependency distance, can be applied to both English and Japanese, the obvious question will be whether the syntactic difficulty of texts in the two languages is the same in terms of dependency distance, despite the striking differences in structure between the two languages. A tentative comparison of spoken texts in English and Japanese will be reported in Chapter 8.
Recent theories of syntactic difficulty

Amongst a number of researchers, some rendering of Chomskyan phrase-structure has been assumed as the basis of a syntactic parsing model. This assumption has led them to make a distinction between top-down and bottom-up parsing strategies. Nevertheless, purely top-down parsing algorithms such as Yngve's (1960) require unbounded memory for head-final structures, although head-initial structures are parsable with finite memory. Similarly, purely bottom-up parsing requires unbounded memory for head-initial structures, in spite of the fact that these algorithms require only finite memory for processing head-final structures. Purely head-initial and -final structures do not cause parsing difficulties for limited human short term memory. It is, therefore presupposed that neither purely top-down nor bottom-up algorithms suffice as plausible human sentence parsing models (Miller & Chomsky, 1963; Gibson, 1991, 1995; Stabler, 1994). In consequence of the inadequacy of both algorithms, some combined algorithms operating with both top-down and bottom up components have been suggested, e.g., a left-corner algorithm (Aho & Ullman, 1972; Kimball, 1973, 1975; Frazier & Fodor, 1978) or a head-corner algorithm (Proudian & Pollard, 1985; Gibson, 1991).

2.1 Frazier and Rayner (1988)

Frazier and Rayner (1988) assume a version of the left-corner algorithm, by which a nonterminal category is constructed at each parsing state of a word parsed from the bottom up. Furthermore, they present the approximate way to measure syntactic complexity across words of a sentence. They hypothesise the Local Nonterminal Count adopted from Frazier (1985) for an approximation of the complexity bound. This is the total value of nonterminal nodes appearing over three adjacent lexical items, whereby nonterminal nodes comprise any node excluding terminal nodes dominating words. They postulate that S and S' nodes contribute a value of 1.5 units, whereas other nonterminals contribute the complexity of 1 unit. The syntactic complexity of a sentence is measured by the Maximal Local Nonterminal Count, which is the largest Local Nonterminal Count of the sentence.

Frazier and Rayner account for the processing difficulties associated with
sentences with multiply embedded sentencial subjects. An illustration of this is given below. (A full phrase-structure tree is also shown only for (2.1a).)

(2.1)

a. that she who fled the country was rescued from the sinking boat was reported.

b. that she was rescued from the sinking boat was reported eased my mind.

(2.1a)
(2.1a) accommodates a relative clause inside a sentential subject, while (2.1b) contains doubly embedded sentential subject. Although both sentences have an equal overall total of nonterminals, (2.1a) is acceptable whereas (2.1b) is not. Frazier and Rayner claim that (2.1b) being harder to process than (2.1a) is due to the need for a larger number of nonterminal counts locally, at the beginning of the sentence. The maximal local nonterminal counts for (2.1), which are measured at the first three words in each sentence, bear out their claim, indicating 8.0, and 8.5 respectively. It is, therefore alleged that processing overload is caused by a maximum local nonterminal count of more than 8.

By Frazier and Rayner’s counting method, the difference in processing difficulty between sentences with sentential subjects versus sentences with extraposed sentential subjects is also explicable, as illustrated in (2.2).

(2.2)

a. That she was rescued from the sinking boat heartened

<table>
<thead>
<tr>
<th>nonterminals</th>
<th>$I_s$</th>
<th>$I_{np}$</th>
<th>$I_{vp}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>processing weight</td>
<td>3.0</td>
<td>2.5</td>
<td>1.0</td>
</tr>
<tr>
<td>me.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

b. It heartened me that she was rescued from the sinking boat.

<table>
<thead>
<tr>
<th>nonterminals</th>
<th>$I_s$</th>
<th>$I_{np}$</th>
<th>$I_{vp}$</th>
<th>$I_{np}$</th>
<th>$I_{vp}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>processing weight</td>
<td>2.5</td>
<td>1.0</td>
<td>1.0</td>
<td>1.5</td>
<td>2.5</td>
</tr>
</tbody>
</table>

The prediction that (2.2a) takes longer to process than (2.2b) is correctly made by the maximal nonterminal counts. 6.5 units is the maximal nonterminal count for the sentential subject sentence (2.2a) which occurs at the first three words. In contrast to (2.2a), the maximal complexity associated with the extraposed sentencial subject counterpart (2.2b) arises over *that she was* whose nonterminal count is 5.0 units.

Moreover, Frazier and Rayner’s complexity metric seemingly provides the right prediction for the processing difficulty in sentences with self-embedded relative clauses, as (2.3) shows.
The boat the woman escaped from had left the devastatingly bombarded country.

The country the boat the woman escaped from had left was devastated.

The sentence with a single self-embedded relative clause such as (2.3a) is easier to process than a double self-embedding like (2.3b). This fact is again in accordance with the maximal nonterminal counts associated with (2.3). The maximal nonterminal count of the singly embedded (2.3a) is 7.5 units, which are observed over the first three words of the sentence. On the other hand, the doubly self-embedded version (2.3b) marks its maximal nonterminal count of 10.0 units over the word string the boat the, which makes the sentence unprocessable.

The data examined so far support Frazier and Rayner's calculation metric of syntactic complexity. It is, however unworkable with a double self-embedding containing the complementiser, THAT, as Gibson (1991) points out. Take (2.4) as an example.

(2.4) #The country that the boat that the woman abandoned left was devastated.

(2.4) as well as (2.3b) is an unacceptable doubly self-embedded sentence. Yet, according to Frazier and Rayner's metric, (2.4) failed to show that its maximal nonterminal count, which is 5 units, exceed the local computational limit observed in (2.1). Their failure to predict the unacceptability of (2.4) is because of the arbitrary decision to measure processing load across these words.
2.2 Gibson (1991, 1995)

Gibson assumes a head-corner algorithm for the human parser under which the parse initiates from the words of an input string in a bottom-up fashion while at the same time the left hand side category is also built whenever the head of a left corner category is located, so that an incremental step of parsing is made possible. Given this underlying parser algorithm, Gibson presumes that a processing weight is associated with local violations of syntactic filters, i.e. the θ-Criterion (Chomsky, 1981), which states that each NP argument bears one θ-role and each θ-role is assigned to one argument, and the Projection Principle (Chomsky, 1981): lexical requirements must be satisfied at all levels of representation. By directly applying these two principles to his theory of processing, Gibson assumes the following two processing properties: The Property of Thematic Reception and The Property of Lexical Requirement. The former is the property of a confirmed constituent in the input which can receive a θ-role but which is not yet assigned one. The latter is the lexical requirement of a head that is not yet fulfilled but which is obligatory in the current structure. These properties are associated with an equal cost, one $x_\theta$ processing load unit (1x$\theta$ PLU), since the θ-Criterion and Projection Principle derive from the same underlying principle, Full Interpretation (Chomsky, 1986a). Gibson (1991) proposes that four PLUs are the memory limit for the human sentence processor.

The thematic interpretation theory of Gibson correctly predicts that the example (2.5) is impossible to process, resulting in memory overload.

(2.5) #The country that the boat that the woman structure [ip{np} [cp{np}] [ip{np} [cp{np}] [ip{np}]
            PLU 1X$\theta$ 1X$\theta$ 1X$\theta$ 1X$\theta$ 1X$\theta$

abandoned left was devastated.

The maximal cost associated with (2.5) is 5X$\theta$ PLUs at the point of maximal complexity, where the NP the woman is being processed. Because 4X$\theta$ PLUs is the maximal bearable memory load for the human processor to execute syntactic processing, (2.5) induces processing overload. The structure of (2.5) is shown below.
The country that the boat that the woman abandoned left was devastated.
Gibson claims that his theory is also accountable for Japanese data of unacceptable self-embedded sentences. The following examples are taken from Gibson (1991).

(2.6)\(^1\)

a. [Taro-ga [Hajime-ga Akira-o suki da] to omotte iru]

\[
\begin{array}{ccc}
\text{sbj} & \text{obj} & \text{likable cop that thinks} \\
1X_{b} & 1X_{b} & 1X_{b}
\end{array}
\]

(Taro thinks that Hajime likes Akira.)

b. ![Taro-ga [Hanako-ga [Hajime-ga Akira-o suki da] to shinjite iru] to omotte iru]

\[
\begin{array}{cccc}
\text{sbj} & \text{sbj} & \text{obj} & \text{believes} \\
1X_{b} & 1X_{b} & 1X_{b} & 1X_{b}
\end{array}
\]

(Taro thinks that Hanako believes that Hajime likes Akira.)

(2.6a) and (2.6b) are a singly and a doubly embedded sentences respectively. For both examples, complementiser to -clauses are self-embedded, since a verb is placed finally in a Japanese sentence. Gibson (1995) claims that a structurally Cased NP is liable to the Property of Thematic Reception, while an inherently Cased NP is not subject to this principle. Hence, according to his assumption, nominative (ga) and accusative (o) (and probably genitive (no)) case-markers, which are actualisations of abstract structural Case, are associated with processing cost 1X\(_{n}\), on the other hand any other case-markers are dissociated from processing load since they are actualisations of inherent Case. The maximal processing cost associated with (2.6a) is, therefore 3X\(_{n}\)PLUs, which occurs while processing the third NP Akira-o. The thematic theory correctly predicts the processability of (2.6a), considering that its maximum load is under the limit of four local 0-Criterion violations.

\(^{1}\)It should be noted that the sentences in (2.6) are marked in two senses. Firstly, in Japanese the default marker of the actor or the experiencer of a NP in an outmost clause is not the subject marker ga but the topic marker wa. Secondly, the embedded sentence Hajime-ga Akira-o suki da is, strictly speaking, ungrammatical, although it is frequently observed in spoken Japanese. Because suki is the stem of an adjective suki-na (likable), Hajime and Akira should be marked by wa and ga respectively, i.e., Hajime wa Akira ga suki da (Talking about Hajime, Akira is likable).
(2.6b)

\[
\begin{align*}
\text{ Category } X_1 & \quad \text{CP} \\
\text{IP} & \quad \text{C'} \\
\text{I'} & \quad \text{C} \\
\text{VP} & \quad \text{C'} \\
\text{V'} & \quad \text{C} \\
\end{align*}
\]

\[
\begin{align*}
\text{ Category } X_2 & \quad \text{CT} \\
\text{IP} & \quad \text{C'} \\
\text{VP} & \quad \text{C'} \\
\text{V'} & \quad \text{V'} \\
\end{align*}
\]

\[
\begin{align*}
\text{PP} & \quad \text{PP} \\
\text{NP} & \quad \text{NP} \\
\text{NP} & \quad \text{NP} \\
\text{NP} & \quad \text{NP} \\
\text{NP} & \quad \text{NP} \\
\text{NP} & \quad \text{NP} \\
\text{NP} & \quad \text{NP} \\
\text{NP} & \quad \text{NP} \\
\end{align*}
\]

\[
\begin{align*}
\text{N'} & \quad \text{P'} \\
\text{N'} & \quad \text{N'} \\
\text{N'} & \quad \text{N'} \\
\text{N'} & \quad \text{N'} \\
\text{N'} & \quad \text{N'} \\
\text{N'} & \quad \text{N'} \\
\text{N'} & \quad \text{N'} \\
\text{N'} & \quad \text{N'} \\
\end{align*}
\]

\[
\begin{align*}
\text{P'} & \quad \text{P'} \\
\text{P'} & \quad \text{P'} \\
\text{P'} & \quad \text{P'} \\
\text{P'} & \quad \text{P'} \\
\text{P'} & \quad \text{P'} \\
\text{P'} & \quad \text{P'} \\
\text{P'} & \quad \text{P'} \\
\text{P'} & \quad \text{P'} \\
\end{align*}
\]

\[
\begin{align*}
\text{N} & \quad \text{P} \\
\text{N} & \quad \text{P} \\
\text{N} & \quad \text{P} \\
\text{N} & \quad \text{P} \\
\text{N} & \quad \text{P} \\
\text{N} & \quad \text{P} \\
\text{N} & \quad \text{P} \\
\text{N} & \quad \text{P} \\
\end{align*}
\]

\[
\begin{align*}
\text{P} & \quad \text{I} \\
\text{P} & \quad \text{I} \\
\text{P} & \quad \text{I} \\
\text{P} & \quad \text{I} \\
\text{P} & \quad \text{I} \\
\text{P} & \quad \text{I} \\
\text{P} & \quad \text{I} \\
\text{P} & \quad \text{I} \\
\end{align*}
\]

\[
\begin{align*}
\text{A} & \quad \text{A'} \\
\text{AP} & \quad \text{V'} \\
\text{V'} & \quad \text{V'} \\
\text{V'} & \quad \text{V'} \\
\text{V'} & \quad \text{V'} \\
\text{V'} & \quad \text{V'} \\
\text{V'} & \quad \text{V'} \\
\text{V'} & \quad \text{V'} \\
\end{align*}
\]

\[
\begin{align*}
\text{C} & \quad \text{C'} \\
\text{I} & \quad \text{C'} \\
\text{V} & \quad \text{C'} \\
\text{V} & \quad \text{C'} \\
\end{align*}
\]

Taro ga Hanako ga Hajime ga Akira o suki da to φ shinjite-iru to φ omotte-iru.
As for (2.6b) shown in the tree structure above, in processing Akira-o there are four NP arguments yet to be assigned 6-roles, which amount to 4xₐ PLUs. This cost is allowable for the parser to operate processing.

In order to account for the unacceptability of (2.6b), Gibson (1995) introduces the Property of self-embedding interference. This property defines the association of a load of 1xₑ PLUs with a predicted category X₁ whose head has not yet been confirmed and which is embedded within another larger predicted category X₂ whose head has not yet appeared either, provided the extended projection features of X₁ are identical to or a subset of those of X₂. For the sake of simplicity of the load calculations, Gibson assumes that xₑ= 1xₑ In the case of (2.6b), the inner embedded IP unlinked with its head is self-embedded inside another such foreseen IP. Consequently the maximal complexity associated with (2.6b) adds up to 4xₑ + 1xₑ = 5xₑ PLUs, which is beyond the limit of short-term memory capacity. Again, Gibson's theory accurately account for processing difficulty with (2.6b).

The examples in (2.7) are given to elucidate further how the thematic interpretation theory manipulates a non-overt operator and an inherently Cased NP for calculations of memory load. These examples are also obtained from Gibson (1995).

(2.7)

a. [Taro-ga [Akira-ga [O Hanako-ga t, suki na] hito-ni hon-o watashi ta] to
   sbj sbj op sbj likable person to book obj handed
   PLU 1xₑ 1xₑ 1xₑ 1xₑ
   it ta.
   said
   (Taro said that Akira handed a book to the person that Hanako likes.)

   sbj to sbj to obj introduced
   PLU 1xₑ 1xₑ 1xₑ
   (Taro said to Hajime that Akira introduced Shigeru to Hanako.)

The maximal memory load required to process (2.7a) is 4xₑ PLUs, which occurs at the point of processing the third overt NP Hanako-ga. Up to this point, there appear three NPs marked by subject particle ga, for which thematic
roles are obligatory, plus one covert operator requiring a \( \theta \)-role. Hence, the sum of the processing cost is 4\( x_q \) PLUs. The self-embedded cost 1\( x_{SE} \) is inapplicable to (2.7a), because the features of the deepest centre-embedded phrase CP are not equivalent to those of the outer embedded CP, i.e. the spec CP of the inner embedding is dominated by a NP which is a projection of the non-lexical operator, whereas the comparable position of the outer CP is unfilled. (2.7b), in contrast, contains inherently Cased NPs. The maximal processing cost associated with (2.7b) occurs at the point of processing the NP Shigeru-o. Up until this fifth NP, there appear three structurally Cased NPs unattached with their \( \theta \)-assigners, namely Taro-ga, Akira-ga and Shigeru-o, and two inherent Cased NPs, Hajime-ni and Hanako-ni, which do not contribute to processing memory load. Hence the maximal memory load associated with (2.7b) is 3\( x_q \) PLUs resulting from three local \( \theta \)-violations, and (2.7b) is fairly easy to process. The example validates the theory’s prediction that (2.7b) gives rise to no processing overload.

Gibson’s theory of processing overload so far seems to be capable of accounting for Japanese data by assuming that a structural Cased NP is associated with memory load, whereas an inherent Cased NP is not. There are, however, severe ambiguities in how the theory explains the unacceptability of a centre-embedded sentence such as a modified version of (2.6b), where the subject particle ga of the NP in the main clause is replaced with the topic marker wa.

(2.8)

\[
\text{#[Taro-wa [Hanako-ga [Hajime-ga Akira-o suki da] to shinjite iru] to omotte}
\]

\[
\begin{array}{cccc}
\text{top} & \text{subj} & \text{subj} & \text{obj} \\
(1x_q) & 1x_q 1x_{SE} & 1x_q & 1x_q \\
\text{iru]
\end{array}
\]

(Taro thinks that Hanako believes that Hajime likes Akira.)

(2.8), the topicalised version of (2.6b) is as unacceptable as (2.6b). The way the thematic theory prevents (2.8) from being acceptable is presumably by assigning 1PLU as well to the topic NP Taro-wa, as Gibson states (quoted from Gibson (1995: 24): the first nominative (Spec-IP) NP ... is moved to topic position so that it is in Spec-CP and receives topic Case “wa”. Consequently, the total memory cost of (2.8) reaches 5PLUs at the point of processing the fourth NP Akira-o. Given these assumptions, the theory
makes a correct prediction about the processing overload of (2.8). Nonetheless, what is unclear in Gibson's account is why he treats a topic NP as a Cased NP, contrary to the usual assumptions.

Moreover the topic NP is presumably regarded as a NP bearing structural Case, so that it becomes responsible for one θ-violation. Given that a topic NP counts 1PLU, the example below displays an overprediction of processing difficulty of the thematic interpretation theory. Consider (2.9).

(2.9)

Naomi-wa [boku-ga [kono kyoku-wa arenji-ga kyoku-no yosa-o dainashi-
i ni shi te iru] to it ta] to omot te iru.  

(2) Naomi thinks that I said that the arrangement spoilt the virtue of this song.)

Under the thematic theory, (2.9) should be predicted to be unacceptable, provided that topic NPs are considered as Cased NPs which contribute to the processing cost. At the worst point of processing (2.9), the partial structure would accommodate five unlinked NPs, namely two topic NPs Naomi-wa and kyoku-wa, two subject NPs boku-ga and arenji-ga, and one object NP yosa-o, which would amount to unbearable 5 PLUs for the parser, and yet (2.9) is quite easy to process. Hence, the theory is too strict.

Furthermore, the topic marker wa, the subject marker ga and the object marker o can be replaced with other particles such as the emphatic marker koso, mo (also), sae (even) or sura (even), or can even be omitted. For NPs marked by these particles and those not marked by any particles, which are nevertheless potential θ-role carriers, it remains uncertain how the theory associates memory cost with these NPs.

As Gibson (1995) mentions, Japanese particles can be idiosyncratic. He gives three examples: (i) the supposed accusative case-marker o used as a realisation of inherent Case, (ii) the alleged dative case-marker ni associated with an

32
Experiencer subject, and (iii) the dative case-marker *ni* indicating the Theme direct object, which is labelled as a quirky case-marker. (2.10), (2.11) and (2.12) demonstrate (i), (ii) and (iii) respectively.

(2.10) Makoto-wa kinou kouen-o aruita.

Top yesterday park obj walked

(Makoto walked in the park yesterday.)

(2.11) Makoto-ni tetsugaku-wa muzukashii.

To philosophy top difficult

(Philosophy is difficult for Makoto.)

(2.12) Makoto-wa tenshi-ni mieru.

Top angel as look

(Makoto looks like an angel.)

Gibson suggests that quirky Case should fall within the domain of the Property of Thematic Reception because quirky Case is not assigned in relation with a particular θ-role but is specified in the lexical entry of the Case-assigner. Thus, according to his theory, the total complexity associated with (2.12) is presumably $2X_s$ PLUs. The example (2.13), however serves as a counterexample to the thematic theory.

(2.13) Naomi-ga Makoto-o [hato-ga heiwa-no shinboru-no you-ni Makoto-ga sbj obj dove sbj peace of symbol of state as sbj  

PLU $1X_s \quad 1X_s \quad 1X_s \quad 1X_s$

seijitsusa-no shinboru-no you-ni mieru] to tatoe ta.

Sincerity of symbol of state as look comp compared past

PLU $1X_s$

(= Naomi made a metaphorical statement about Makoto that he looks like a symbol of sincerity as the dove is a symbol of peace.)

Before the parser arrives at the first verb *mieru* in the sentence (2.13), the maximal complexity associated with it reaches $5X_s$ PLUs, given the assumption that quirky Cases as well as structural Cases are affiliated with processing cost. At the point of processing the NP *you-ni* there are five NPs preceding
their θ-assigners, which results in a processing cost which exceeds short-term memory capacity. The example (2.6) therefore, is wrongly predicted to create processing difficulty by the theory, whereas it is actually easy to process. In addition, quirky case *ni selected by the verb *mieru* can also be replaced by *-no you-ni* as in the example (2.13), *ga-gotoku, no-gotoku, -ppoku, mitai-ni, or mitaku*, which all crudely mean "like". The theta-based theory seems to encounter predicaments in justifying applying the Principle of Thematic Reception to a phrase which is not explicitly indicated by straightforward structural Case.

The fundamental difficulty with application of the thematic interpretation theory of Gibson to Japanese data seems to lies in the neglect of the flexible omissibility of the language and the disparity between Japanese particles and case. As reminded in section 1, deletion of any element of a sentence is permissible in Japanese, provided that missed items are retrievable by the addressee. For instance, the conversation below is possible in the situation where Akira and Naomi are going to have dessert.

\[
\begin{align*}
(2.14) \quad & \text{a. Akira: Naomi nani.} \\
& \text{b. Akira: nani taberu.}
\end{align*}
\]

The sentences (2.14a) suffice in order for Naomi and Akira to communicate each other although it is impossible to reconstruct the full versions of the speaker's intended sentences. (2.14a) are also grammatical although the argument NPs are lacking θ-roles because of the absence of their predicates. Further (2.14b) is another possible utterance for Akira conveying the same message as (2.14a). In (2.14b) the verb *taberu* fails to θ-mark its arguments since the argument NP expressing the Agent of the activity of eating is implicit. Observing these allowable sentences violating the θ-Criterion, it seems to be the case that the θ-Criterion is no longer a prerequisite for the Japanese grammar.

On the subject of particles, it has been witnessed in (2.10), (2.11) and (2.12) that the overt realization of abstract case by means of putative morphological case is not necessarily coherent in Japanese. Furthermore, the particles
such as the topic marker wa and the subject marker ga may manifest different semantic interpretations. Consider the examples in (2.15).

(2.15) a. kinou Akira-wa chero-o hii ta.
    yesterday top cello obj played
    (Yesterday Akira played the cello.)

b. kinou Akira-ga chero-o hiita.

(2.15a) is an unmarked sentence for the English counterpart 'Yesterday Akira played the cello.', although the agent of the verb hiku 'to play' is identified by the topic marker wa. In contrast, (2.15b) whose agent NP is adjoined with the subject marker is marked, crudely meaning that yesterday not others but Akira played the cello. In this case, ga is functioning not only as the subject marker but also the focus marker. Given that the nominative case is defined to be an element signifying the external argument of a verb, this term appears unsuited to depict the whole meaning of ga.

2.3 Hawkins (1994)

Recognising the need for a more gradient approach to processing difficulty, Hawkins (1994) offers a general theory of structural complexity on the basis of which he develops a theory of word order, Early Immediate Constituents (EIC). I shall introduce in brief some of his key concepts in terms of syntactic complexity and his method of measuring syntactic difficulty proposed within the framework of these theories.

Although Hawkins agrees on the basic idea of syntactic complexity proposed in Miller and Chomsky (1963) and Frazier (1985) that complexity correlates with the amount of phrase-structure, he argues against their particular formulae and therefore, defines the notion of a Structural Domain and its complexity as follows:

**Structural Domain (SD)**
A Structural Domain consists of a grammatically and/or psycholinguistically significant subset of structurally related nodes in a tree dominated by a constituent C. (Hawkins, 1994: 25)
**Structural Complexity**

The structural complexity of an SD is measured by counting the set of nodes within it: the complexity of SD₁ exceeds that of SD₂ iff SD₁ > SD₂. (Hawkins, 1994: 29)

In elaborating his counting system of syntactic difficulty, he presents a type of Structural Domain, viz. Constituent Recognition Domain (CRD), with which on-line phrase-structure recognition is made possible. This term is defined as follows:

**Constituent Recognition Domain (CRD)**

The CRD for a phrasal mother node M consists of the set of terminal and non-terminal nodes that must be parsed in order to recognise M and all immediate constituents (ICs) of M, proceeding from the terminal node in the parse string that constructs the first IC on the left, to the terminal node that constructs the last IC on the right, and including all intervening terminal nodes and the non-terminal nodes that they construct. (Hawkins, 1994: 58-59)

Considering the fact that the more terminal and non-terminal nodes the CRDs for VP and S contain, the more acceptability levels, ease of processing, and textual frequency diminish, Hawkins poses two procedures for measuring structural complexity: (i) counting the number of all the terminal and non-terminal nodes in a CRD (ii) calculating the ratio between the number of higher nodes, i.e. ICs and that of lower nodes, i.e. non-ICs that need to be processed in order to recognise the ICs in a CRD. Hence, the maximisation of IC-to-non-IC ratios for a given CRD increases the efficiency and ease with which it is recognised. Let us take a Heavy NP Shift construction as an example to see how both the methods are applied.

(2.16a) I gave NP[the valuable book that was extremely difficult to find] PP[to Mary].

(2.16b) I gave PP[to Mary] NP[the valuable book that was extremely difficult to find].

(Hawkins, 1994: 57)
N.B. The CRD for the top VP in (2.16a) is surrounded by a dotted line.

The structural complexity of the CRD for the higher VP

<table>
<thead>
<tr>
<th>method</th>
<th>results</th>
</tr>
</thead>
<tbody>
<tr>
<td>(i)</td>
<td>32 terminal and non-terminal nodes to be scanned</td>
</tr>
<tr>
<td>(ii)</td>
<td>IC/non-IC ratio = 3/28 = .107 (= 10.7%)</td>
</tr>
</tbody>
</table>
The syntactic difficulty for (2.16a) and (2.16b) measured in terms of IC recognition within the CRDs for the higher VPs with the methods (i) and (ii) seems to indicate correctly the relative ease of processing the shifted sentence (2.16b). Hawkins adopts the method (ii), pointing out that, in contrast

3The phrase-structure tree presented here is Hawkins’ version (Hawkins, 1994: 60). As for the case of a heavy NP-shift sentence however, attaching a shifted NP directly under the top S node seems to be a standard analysis, which inevitably yields a different result for the structural complexity of the CRD for the higher VP in (2.16b).

N.B. The CRD for the top VP in (2.16b) is surrounded by a dotted line.

The structural complexity of the CRD for the higher VP

<table>
<thead>
<tr>
<th>method</th>
<th>results</th>
</tr>
</thead>
<tbody>
<tr>
<td>(i)</td>
<td>8 terminal and non-terminal nodes to be scanned</td>
</tr>
<tr>
<td>(ii)</td>
<td>IC/non-IC ratio = 3/8 = .375 (= 37.5%)</td>
</tr>
</tbody>
</table>

(2.16b)
with (i) it enables us to compute syntactic difficulty in a cross-linguistic or structurally sensitive environment.

(ii) is allegedly advantageous when comparing one language with VP node within S to another having no VP inside S (cf. Hale 1983), or when contrasting Extraposition constructions in English. For a language with no VP node in S, there will clearly be no CRD for the VP whereas there will be more ICs contained in the S node, and consequently a larger CRD for the S than that of a language having VP inside of S. Since constituents that the S node of the VP-free language is comprised of are distributed in the VP and S nodes in the language accommodating VP, Hawkins contends that a fair comparison of syntactic difficulty can be made between the two languages by measuring relative complexity in terms of ratios of ICs-to-non-ICs for given corresponding phrases in the languages under consideration.4 It seems however, that both the methods (i) and (ii) ought to give plausible results for a cross-linguistic comparison providing any given phrase in one language and its equivalent in another are taken into consideration.

As for the structurally sensitive conditions such as Extraposition in English, consider these two methods applied to an unextraposed sentence (2.17a) and its extraposed counterpart (2.17b) below.

(2.17a)  \[s_{\text{S}}[\text{That Bill was frightened}]_{\text{VP}}[\text{surprised}]_{\text{NP}}[\text{Mary}]\].

(2.17b)  \[s_{\text{S}}[\text{It}]_{\text{VP}}[\text{surprised}_{\text{NP}}[\text{Mary}]_{\text{S}}[\text{that Bill was frightened}]\].

(Hawkins, 1994: 65)

---

4Although the IC-to-non-IC ratio for a whole sentence is defined as the aggregate of the scores for all CRDs within the sentence (cf. Hawkins' summary quoted below in Calculating IC-to-non-IC ratios), Hawkins actually measures an average of ratios rather than an aggregated value in comparing more than one CRD (cf. Table 2.1).
(2.17a)

N.B. ⊙ = the CRD for VP; ○ = the CRD for S.

The structural complexity of the CRD for (2.17a)

<table>
<thead>
<tr>
<th>method</th>
<th>CRD for VP</th>
<th>CRD for S</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>(i)</td>
<td>6</td>
<td>17</td>
<td>23</td>
</tr>
<tr>
<td>(ii)</td>
<td>2/3 (= 66.6%)</td>
<td>2/14 (= 14.2%)</td>
<td>34/84 (= 40.4%)</td>
</tr>
</tbody>
</table>
NR

Maiy that / NP

V AdjP N

Bill Ad was

frightened

N.B. ⊙ = the CRD for VP; ○ = the CRD for S.

The structural complexity of the CRD for (2.17b)

<table>
<thead>
<tr>
<th>method</th>
<th>CRD for VP</th>
<th>CRD for S</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>(i)</td>
<td>9</td>
<td>7</td>
<td>16</td>
</tr>
<tr>
<td>(ii)</td>
<td>3/5 (= 60.0%)</td>
<td>2/4 (= 50.0%)</td>
<td>22/40 (= 55.0%)</td>
</tr>
</tbody>
</table>

Hawkins claims that a sensitive treatment is required when we make a comparison of (2.17a) to its extraposed counterpart (2.17b), since the size of the CRDs for the two higher VPs in (2.17) differs due to their structural difference; according to Hawkins' analyses, within the VP in (2.17a) two ICs are recognised, namely the verb surprised and the NP Mary, while three ICs, V, NP and S' are attached to the mother node VP in (2.17b). Both the methods (i) and (ii) give the implausible result in terms of processing difficulty that the CRD size for VP in (2.17a) is smaller than that in (2.17b). Thus, he suggests that the average complexity values of the CRDs for the VP and the S must be considered, which results in the correct prediction that the extraposed sentence (2.17b) is relatively easier to process than the unextraposed version (2.17a) by means of both methods (i) and (ii). It appears that both the methods are equally workable for measuring syntactic
difficulty, so there is no reason to prefer (ii) rather than (i). To simplify the analysis, Hawkins proposes a simplified version of (ii) which uses the ratio of the number of ICs to that of words instead of non-ICs within a given CRD, focusing on the fact that an increase of words signifies more additional structure, which in turn means more non-ICs. Table 2.1 shows the comparison of IC-to-non-IC and IC-to-word ratios of (2.16) and (2.17).

Table 2.1 IC-to-non-IC and IC-to-word ratios of (2.16) and (2.17)

<table>
<thead>
<tr>
<th>Sentence</th>
<th>IC-to-non-IC ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CDR for VP</td>
</tr>
<tr>
<td>(2.16a)</td>
<td>3/28 (10.7%)</td>
</tr>
<tr>
<td>(2.16b)</td>
<td>3/8 (37.5%)</td>
</tr>
<tr>
<td>(2.17a)</td>
<td>2/3 (66.6%)</td>
</tr>
<tr>
<td>(2.17b)</td>
<td>3/5 (60.0%)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sentence</th>
<th>IC-to-word ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CDR for VP</td>
</tr>
<tr>
<td>(2.16a)</td>
<td>3/11 (27.3%)</td>
</tr>
<tr>
<td>(2.16b)</td>
<td>3/4 (75.0%)</td>
</tr>
<tr>
<td>(2.17a)</td>
<td>2/2 (100%)</td>
</tr>
<tr>
<td>(2.17b)</td>
<td>3/3 (100%)</td>
</tr>
</tbody>
</table>

Juxtaposing the two different ratios in Table 2.1, we can observe that in general they demonstrate roughly the same relative ranking of complexity, but that the ranking varies in the figures for the CRD for VP in (2.17a) and (2.17b), and also that for the CRD for S in (2.17b); these three values vary in the IC-to-non-IC ratio on the one hand and they are identical in the IC-to-word ratio on the other. Although the IC-to-word ratios for the overall complexity of (2.17) exhibit the correct relative difference between (2.17a) and (2.17b), which is analogous to that of the IC-to-non-IC ratios, in the strict sense IC-to-word ratios do not reproduce exactly the same relative results gained by IC-to-non-IC ratios. In short, Hawkins' simplified measure is NOT equivalent to the IC-to-non-IC ratio. However, Hawkins summarises his way of quantifying syntactic complexity as follows:
**Calculating IC-to-non-IC ratios**

The IC-to-non-IC ratio for a CRD is calculated by dividing the number of ICs in the domain by the total number of non-ICs (or words alone) in that domain, expressing the result as a percentage. The ratio for a whole sentence is the average of the scores for all CRDs within the sentence.⁵ (Hawkins, 1994: 76-77)

However, whichever method of counting is adopted, Hawkins’ theory has serious weakness. His counting technique, as Hawkins himself observes, faces the theoretical problem that the constituent structure of a given phrase depends profoundly on the analyses that various theories entail. For instance, returning to the example (2.17b) of English extraposition, we can see that two competing analyses on the S' node are possible because it can be treated as either an IC of the VP, which is Hawkins’ version of analysis as shown in (2.17b), or that of the top node S. Evidently these different structurings of the sentence result in very different results for its syntactic complexity. The alternative phrase-structure analysis to (2.17b) and its results of IC-to-word ratios are given below.

⁵The word 'aggregate' rather than 'average' is used in the definition of Calculating IC-to-non-IC ratios.
The structural complexity of the CRDs for (2.17a), (2.17b) and (2.17b')

<table>
<thead>
<tr>
<th>Sentence</th>
<th>IC-to-word ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CDR for VP</td>
</tr>
<tr>
<td>(2.17a)</td>
<td>2/2 (100%)</td>
</tr>
<tr>
<td>(2.17b)</td>
<td>3/3 (100%)</td>
</tr>
<tr>
<td>(2.17b')</td>
<td>2/2 (100%)</td>
</tr>
</tbody>
</table>

It is true that (2.17b') still shows the expected simplifying effect of extraposition, but the IC-to-word ratios for the whole of (2.17b) and (2.17b') indicate that there is a marked contrast between them, with (2.17b') retaining a more complex structure than (2.17b). The fact that two different phrase-structures assigned to an identical sentence demonstrate differing syntactic complexity raises serious questions for Hawkins' system for measuring complexity.

2.4 Gibson (1998)

Gibson (1998) proposes a new theory to reflect the earlier theory reported in Section 2.2 - the Syntactic Prediction Locality Theory (SPLT) concerning the relationship between the language processing mechanism and the memory resources available for processing. This theory of the linguistic computational resources, which is aimed at accounting for a large array of processing phenomena, has two components: an integration resource component and a memory resource component. Both of these components are hypothesised to be affected by locality, the distance measured in terms of the number of new discourse referents in the intervening syntactic structure. I shall explain the nature of the two resource components, integration and memory cost, the relationship between the two within the theory, and how syntactic difficulty is measured, using as an example the lower complexity of subject-extracted relative clauses in comparison with object-extracted relative clauses.

Syntactic processing necessitates integration whereby a new linguistic input is incorporated into the current syntactic structure under processing. Gibson assumes that linguistic integration involves a consistent quantity of computational resources and additional computational resources varying
according to the distance between a new input word and the word to which the input is attached, corresponding to the resources required to reactivate the non-local element. The amount of these two resources is presumed to be determined by a linear integration cost function $I(n)$ whose intercept is zero, namely $I(n) = K_{int} \cdot n$, where $n$ = the number of new discourse referents between the current word and its target in the parsed structures, and $K_{int} = 1$, the slope of the line. Concerning the characteristics of $n$ in $I(n)$, Gibson postulates that integration cost increments are induced when processing words that contain new discourse referents, which are potential referents for a later anaphoric expression. The main idea in this assumption is that constructing a structure for a new discourse referent involves considerable computational resources, and integrating the current structure to older structures expends more resources as they have weaker activation or have lost it. Integration cost is calculated in terms of energy unit (EUs). The definition of integration cost that the SPLT hypothesises is as follows:

**Linguistic Integration Cost**

The integration cost associated with integrating a new input head $h_2$ with a head $h_1$ that is part of the current structure for the input consists of two parts: (1) a cost dependent on the complexity of the integration (e.g. constructing a new discourse referent); plus (2) a distance-based cost: a monotone increasing function $I(n)$ energy units (EUs) of the number of new discourse referents that have been processed since $h_1$ was last highly activated. For simplicity, it is assumed that $I(n) = n$ EUs.

(Gibson, 1998: 12-13)

The second component of the SPLT is memory cost associated with retaining unintegrated syntactic categories such as a head noun for the subject and a head verb for the predicate. In order to specify the inventory of syntactic categories, Gibson assumes a syntactic theory minimising the number of functional categories, i.e. Head-driven Phrase Structure Grammar (Pollard & Sag, 1994) or Lexical Functional Grammar (Bresnan, 1982). Gibson also

For simplicity of the discussion, Gibson assumes a linear function, but he points out that this assumption is an oversimplification, since empirical evidence shows that the integration function is nonlinear in the limit (cf. Church, 1980, Gibson et al. 1996) Also, alternatively it is assumed that the integration cost function is linear with a zero intercept: $I(n) = K_{int} \cdot n$, where $K_{int}$ is constant (Gibson, 1998:30).
hypothesises that additional memory load is incurred for remembering other syntactic heads that words require to compose a syntactically well-formed sentence, e.g. an empty category NP.

It is proposed relating to the locality-based memory cost that the prediction of the matrix predicate is cost-free on the grounds that a predicate is always anticipated by the parser, so the parsing algorithm predicts a predicate, which is free of memory cost. Gibson admits that this is an oversimplification, because the hypothesis predicts a substantial complexity contrast between (2.18a) with a construction where the matrix verb in the matrix is associated with no memory cost and (2.18b) having the same construction in (2.18a) as an embedded clause, hence an increasing memory cost is incurred for it.

(2.18a) The reporter that the senator attacked ignored the president.

(2.18b) The editor said that the reporter that the senator attacked ignored the president.

(Gibson, 1998: 27)

The prediction of the matrix verb ignored in (2.18a) costs no memory load during the processing of (2.18a) since the verb is in the matrix clause. By contrast, the prediction of the embedded verb of the matrix clause in (2.18b), also ignored, by virtue of being positioned in the embedded clause is associated with a memory cost during processing (2.18b). It follows that processing (2.18b) should consume more memory resources than (2.18a) at the point of processing the subject noun senator in the relative clause where the maximal memory load is required. This prediction however, is incorrect; (2.18a) and (2.18b) ought to be the same in their syntactic difficulty. To account for the lack of difference in syntactic difficulty between (2.18a) and (2.18b), Gibson further proposes the following principle.

THE CLAUSE-BASED CLOSURE PRINCIPLE

The initiation of a new clause causes closure of an earlier clause whose preferred dependents have been satisfied.

(Gibson, 1998: 29)

The implementation of this principle means that the parser no longer has to consider some potential attachment sites which have been removed by a
closure as potential attachment sites. Experimental evidence for this principle is provided by many researchers (Marslen-Wilson, Tylor & Seidenberg, 1978; Frazier & Fodor, 1978; Gibson et al., 1996). Furthermore, it is assumed that not only the predicted matrix predicate but any top-level predicted predicate in the current memory buffer is associated with no memory cost. This assumption together with the principle above is able to explain that there is no difference in syntactic difficulty between (2.18a) and (2.18b). At the point of processing *that* in (2.18b), the parser thinks that the top-level predicted predicate will be the embedded verbal complement of the matrix clause, if there is one, because the matrix verb *said* has been shunted away from the working memory buffer. Consequently, in processing the word *ignored*, no memory cost is associated with it, for this time it is the highest predicted predicate whose memory cost is free, which results in the equal difficulty of (2.18a) and (2.18b). First Gibson defines the term matrix argument clause and formalises the refinement of the zero-cost predicted hypothesis as follows:

**Matrix argument clause**

A matrix argument clause is either the matrix clause or an argument clause of a matrix argument clause.

(Gibson, 1998: 29)

**The zero-cost top-level predicate hypothesis**

The prediction of the matrix argument predicate, \( V_0 \), is associated with no memory cost.\(^7\)

(Gibson, 1998: 29)

It is hypothesised for simplicity that like the integration cost function, the memory cost function is a discourse-based locality function and linear with a zero intercept, i.e. \( M(n) = K_{\text{Mem}} \ast n \), where \( K_{\text{Mem}} = 1.\)\(^8\) The reasoning that

\(^7\)Although the expression 'the top-level matrix argument predicate' is actually used in this hypothesis, it does not necessarily have to be top-level, since by definition a matrix argument clause can also be an argument clause of a matrix argument clause.

\(^8\)The evidence (Church, 1980; Gibson et al. 1996) suggests that more accurately the memory cost function is similar to the sigmoid function which increases monotonically and heads asymptotically towards a maximal complexity. However, Gibson assumes the linear version of the function since all the results he reported in the paper are consistent with the sigmoid function, and easier to understand. For the memory cost function as well, the alternative assumption is given that \( I(n) = K_{\text{mem}} \ast n \), where \( K_{\text{mem}} \) is
prompts this assumption is that lexical materials expend memory resources in order to build their category predictions when they are parsed, so each new discourse referent causes a memory cost. Memory cost is counted in terms of memory units (MUs). The summary of the SPLT memory cost hypothesis is given below.

**Syntactic Prediction Memory Cost**

(a) The prediction of the matrix argument predicate, $V_o$, is associated with no memory cost.

(b) For each required syntactic head $C_i$ other than $V_o$, associate a memory cost of $M(n)$ memory units MUs where $M(n)$ is a monotone increasing function and $n$ is the number of new discourse referents that have been processed since $C_i$ was initially predicted.

(Gibson, 1998: 15, 29)

The underlying parser of the SPLT must be able to construct structures that contain information on syntactic categories, since new discourse referents required to build a sentence are assumed to induce memory cost. The requirement that the parser ought to retain a connected sentence structure representation for an input string after parsing each word suggests that the theory underlies a lexically-driven predictive parser. Following Gibson (1991), it is presupposed in particular that each lexical entry for a word $w$ in the parser comprises information on a phrase structure for $w$ plus potential syntactic structures that can immediately follow $w$ in order to form a completely connected grammatical sentence structure. Thus, the procedure for building syntactic structure by the parser is (i) looking up the input word $w$, in the grammar to compile into the parser the information about its syntactic structure along with structures for the syntactic categories of the words that can immediately follow $w$, (ii) matching the categories of the lexical entries to the predictions in the parsed structures constructed thus far, and (iii) transferring a connected syntactic representation onto the next parse state until the parse is completed.

The SPLT incorporates the hypothesis of the single pool of working memory resources by Just and Carpenter (1992), thus linguistic storage and integration consume the same memory resources. Given a time required to constant (Gibson, 1998:30).
perform a linguistic integration, the conversion formula between an amount of energy necessary to perform an integration and an amount of memory cost is as follows:

**THE RELATIONSHIP BETWEEN MEMORY AND INTEGRATION COST**

An energy unit (EU) = memory unit (MU) * time unit (TU)

(Gibson, 1998: 15)

From these assumptions it follows that the relationship between consumed storage space and required integration time is not linear but a reversed relationship; if there is more storage space used up (a smaller working memory capacity), then integration proceeds more slowly. This non-linear relationship is formally stated below.

**THE TIMING OF LINGUISTIC INTEGRATION**

\[ t_{\text{struct-integ}} = C \times I_{\text{struct-integ}} \left( M_{\text{capacity}} - M_{\text{current-memory-used}} \right); \]

where

- \( t_{\text{struct-integ}} \) is the time required to perform an integration;
- \( C \) is a constant
- \( I_{\text{struct-integ}} \) is the quantity of energy resources (in EUs) required to perform the integration as determined by the function \( I(n) \) in linguistic memory cost
- \( M_{\text{capacity}} \) is the linguistic working memory resources capacity of the listener/reader (in MUs)
- \( M_{\text{current-memory-used}} \) is the memory resources already in use (in MUs) as determined by the function \( M(n) \) in syntactic prediction memory cost.

(Gibson, 1998: 16)

According to King & Just (1991), in their experiments on individual differences in working memory capacity and reading time, subjects with small memory capacity, which was measured in terms of the reading span task (Daneman & Carpenter, 1980), read more slowly than subjects with larger memory capacity overall and also at points where high memory cost had been predicted. The results prompted the single memory pool hypothesis that the same working memory resources were accessed for linguistic retention and integration (Just & Carpenter, 1992). Also, the timing of linguistic integration is justified by the results, in particular with smaller working memory space available, linguistic computation takes a longer
amount of time.

Next, exploring a well-known phenomenon of relative syntactic difficulty between an object-extracted relative clause and a subject-extracted relative clause in English, we shall observe how the integration and memory cost assumed by the SPLT makes a correct prediction as to their relative difficulty. Consider the following examples (2.19) accompanied by their tree diagrams.

(2.19a) \[_s \text{The reporter [}_s \text{ who [}_s \text{ the senator attacked}]] \text{ admitted the error}.\] (Gibson, 1998: 2)

(2.19b) \[_s \text{The reporter [}_s \text{ who [}_s \text{ attacked the senator}]] \text{ admitted the error}.\] (Gibson, 1998: 2)
In the minimal pair of sentences (2.19), the sentence with the object-extracted relative clause, (2.19a) is affirmed to be more difficult to process than the subject-extracted relative clause version, (2.19b). This is confirmed by empirical evidence such as eye fixation patterns (Holmes & O'Regan, 1981), reading times (King & Just, 1991; Caplan & Waters, 1996a), and also a study of the volume of blood flow in the brain (Just et al., 1996), which showed that the blood flow in language areas in the brain was more active in processing object-extractions than subject-extractions.

The source of the relative difficulty is due neither to lexical differences since both the sentences are constructed with exactly the same words, nor a "garden-path" effect because there are no serious ambiguity points in (2.19). Thus, the only possible reason for the relative difficulty in (2.19) should be the difference in the measure of memory resources required for processing. Gibson assumes that the relative syntactic difficulty of a sentence is determined by the total quantity of memory resources which are expended at any point of the parse of the sentence (Gibson, 1991). Hence, a sentence is assessed to be more syntactically difficult if more memory resources are required at the point of maximal memory load. In order to examine the predictions for the difficulty of (2.19) by the SPLT, we need to consider how much of the memory resources is used for integration and for linguistic memory storage at each parsing state, because the SPLT assumes a single memory resource pool hypothesis. First, let us consult the SPLT integration complexity profiles for (2.19) given below. The profiles below are based on Gibson (1998: 20-21).

(2.20a) Integration cost for (2.19a)

<table>
<thead>
<tr>
<th>The</th>
<th>reporter</th>
<th>who</th>
<th>the</th>
<th>senator</th>
<th>attacked</th>
<th>admitted</th>
<th>the</th>
<th>error.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(IN EUs)</td>
<td>-</td>
<td>I(0)</td>
<td>I(0)</td>
<td>I(0)</td>
<td>I(1)+I(2)</td>
<td>I(3)</td>
<td>I(0)</td>
<td>I(0)+I(1)</td>
</tr>
<tr>
<td>(*K_{int} EUs)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Maximal complexity: 3*K_{int} EUs in the region from attacked to admitted
Total complexity: 7*K_{int} EUs
(2.20b) **Integration cost for (2.19b)**

<table>
<thead>
<tr>
<th></th>
<th>The</th>
<th>reporter</th>
<th>who</th>
<th>attacked</th>
<th>the</th>
<th>senator</th>
<th>admitted</th>
<th>the</th>
<th>error.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(IN EUs)</td>
<td>-</td>
<td>I(0)</td>
<td>I(0)</td>
<td>I(0)+I(1)</td>
<td>I(0)</td>
<td>I(0)+I(1)</td>
<td>I(3)</td>
<td>I(0)</td>
<td>I(0)+I(1)</td>
</tr>
<tr>
<td>(*K_{int} EUs)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Maximal complexity: 3*K_{int} EUs at the point of admitted
Total complexity: 6*K_{int} EUs

N.B. The rows of (*K_{int} EUs) show the actual values of integration cost required for integration at each point of processing a word. Each numeral should be multiplied by a constant K_{int} in that I(n) = K_{int} *n.

In (2.20a), no integration takes place for the first input word *The*. At the second word, an integration of the word *reporter* with the preceding word *The* occurs, which costs I(0) EUs because no new discourse referents are involved between the determiner *the* and the current input word *reporter*. On top of that, the cost for constructing a new discourse referent "the reporter" is initially ignored, since it is only counted as part of distance-based element of integration cost. The ignoring of this cost applies to other words that contain new discourse referents: the nouns "reporter", "senator" and "error", and the tensed verbs "attacked" and "admited". Next, linking the following word *who* to the most recent word *reporter* again leads to an integration cost of I(0) EUs owing the lack of new discourse referents since "reporter" was processed where the referent of *who* is attached. The following word *the* is incorporated into the parsed structure at no integration cost since no new discourse referent is involved. The next word *senator* requires I(0) EUs for its integration with *the* as there is no intervening new discourse referents. Two integrations arise for processing the next input *attacked*. The first integration involves attaching the input to its subject NP *the senator*, which results in a cost of I(1) EUs because one new discourse referent "attacked" has been processed since the subject NP *the senator* was processed. The second integration comprises linking the input with an empty category as its object and coindexing it with the relative pronoun *who*, whereby an additional I(2) EUs is consumed because of an involvement of two new discourse referents since the target word *who* was processed: the object referent "the senator" and the event referent.
“attacked”. The integration of the next verb admitted to the matrix subject NP the reporter induces I(3) EUs because there are three new discourse referents since the head NP the reporter was parsed: the object referent “the senator” and the two event referents “attacked” and “admitted”. An integration costing I(0) EUs takes place at the next word the, since no new discourse referent is involved in the linking. Finally, the last word “error” is integrated into the preceding determiner the at a cost of I(0) EUs, and the integration of the NP the error into the verb admitted as its object costs I(1) EUs thanks to one new discourse referent “error” that has been processed since its head admitted was parsed.

As for (2.20b), the integration cost profile is the same as the one for its object-extracted version presented in (2.20a), except the region of the embedded clause, attacked the senator. The integration cost for the verb attacked in the subject-extraction is I(0)+I(1) EUs, with which the verb is integrated into an empty category in its subject position and the gap is coindexed with the relative pronoun who. The attachment involves the intervention of one new discourse referent “attacked”. The next word the is integrated at an integration cost of I(0), corresponding to the linking of the input with the embedded verb attacked where no new discourse referent participates. The next input senator involves two integrations: the noun senator integrating with the determiner the at a cost of I(0) EUs, and the NP the senator integrating with the verb attacked at a cost of I(1) EUs, since one new discourse referent “the senator” is involved since the head verb attacked was input.

In terms of the maximal complexity measured by integration cost, both (2.20a) and (2.20b) mark 3K_int EUs, so the syntactic complexity of (2.19a) and (2.19b) is the same so far as integration cost is concerned. However, the total complexity for these sentences suggests correctly the relative syntactic difficulty of (2.19a), since the total complexity of (2.19a), 7K_int EUs, is greater than that of (2.19b), namely 6K_int EUs. This is mainly because the maximal complexity of 3K_int EUs is incurred over the two words attacked and admitted in (2.20a), while it is associated with only the matrix verb admitted in (2.20b).

The other component in the SPLT that accounts for syntactic difficulty contrasts, viz. memory cost is considered in the following. With the aid of
memory cost profiles, we will investigate whether the SPLT makes a plausible prediction about the difference in syntactic difficulty between the object-extracted relative clause sentence (2.19a) and the object-extracted counterpart (2.19b). The memory complexity profiles for (2.19a) and (2.19b) are displayed below in (2.21a) and (2.21b) respectively. The profiles below are based on Gibson (1998: 33-34).

(2.21a) Memory cost for (2.19a)

<table>
<thead>
<tr>
<th>SYNTACTIC PREDICTION</th>
<th>The</th>
<th>reporter</th>
<th>who</th>
<th>the</th>
<th>senator</th>
<th>attacked</th>
<th>admitted</th>
<th>the</th>
<th>error.</th>
</tr>
</thead>
<tbody>
<tr>
<td>matrix verb</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>matrix subject</td>
<td>M(0)</td>
<td>*</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>matrix object NP</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>M(0)</td>
<td>M(0)</td>
<td>*</td>
</tr>
<tr>
<td>embedded subject NP</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>M(0)</td>
<td>*</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>embedded verb</td>
<td>-</td>
<td>-</td>
<td>M(0)</td>
<td>M(0)</td>
<td>M(1)</td>
<td>*</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>wh-pronoun gap</td>
<td>-</td>
<td>-</td>
<td>M(0)</td>
<td>M(0)</td>
<td>M(0)</td>
<td>*</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>MEMORY COST (MUs)</td>
<td>M(0)</td>
<td>0</td>
<td>2M(0)</td>
<td>3M(0)</td>
<td>2M(1)</td>
<td>0</td>
<td>M(0)</td>
<td>M(0)</td>
<td>0</td>
</tr>
<tr>
<td>(*K&lt;sub&gt;Mem&lt;/sub&gt; MUs)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Maximal complexity: 2*K<sub>Mem</sub> MUs at the point of processing the embedded NP subject senator
Total complexity: 2*K<sub>Mem</sub> MUs

(2.21b) Memory cost for (2.19b)

<table>
<thead>
<tr>
<th>SYNTACTIC PREDICTION</th>
<th>The</th>
<th>reporter</th>
<th>who</th>
<th>attacked</th>
<th>the</th>
<th>senator</th>
<th>admitted</th>
<th>the</th>
<th>error.</th>
</tr>
</thead>
<tbody>
<tr>
<td>matrix verb</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>matrix subject</td>
<td>M(0)</td>
<td>*</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>matrix object NP</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>M(0)</td>
<td>M(0)</td>
<td>*</td>
</tr>
<tr>
<td>embedded subject NP</td>
<td>-</td>
<td>-</td>
<td>*</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>embedded verb</td>
<td>-</td>
<td>-</td>
<td>M(0)</td>
<td>*</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>wh-pronoun gap</td>
<td>-</td>
<td>-</td>
<td>M(0)</td>
<td>M(0)</td>
<td>M(0)</td>
<td>*</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>MEMORY COST (MUs)</td>
<td>M(0)</td>
<td>0</td>
<td>2M(0)</td>
<td>M(0)</td>
<td>M(0)</td>
<td>0</td>
<td>M(0)</td>
<td>M(0)</td>
<td>0</td>
</tr>
<tr>
<td>(*K&lt;sub&gt;Mem&lt;/sub&gt; MUs)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Maximal complexity: 0*K<sub>Mem</sub> MUs
Total complexity: 0*K<sub>Mem</sub> MUs
N.B. • The left-most column in each table shows all the items that require syntactic predictions during the parse.  
• The memory load associated with the syntactic prediction is indicated in a cell of the word position in which the cost arises, in terms of the function $M(n)$ where $n$ is the number of new discourse referents that have been processed since the prediction was first made.  
• An asterisk is entered in a position where a syntactic prediction is satisfied.  
• A dash is placed in a position corresponding to syntactic predictions that are inactive.  
• The number of words giving rise to memory cost increments are three, i.e. the head noun of the matrix subject NP, reporter, the head noun of the embedded NP, senator, and the verb of the embedded VP, attacked.  
• The prediction of the matrix verb admitted is associated with no memory cost on account of the zero-cost top-level predicate hypothesis, so zero is presented in its prediction row.  
• The total memory cost for each word is shown in the MEMORY COST (MUs) row in terms of the function $M(n)$.  
• The last row of $^{*}K_{\text{Mem}}$ MUs demonstrates the actual values of memory cost required for storage at each indexed word position. Each numeral should be quantified by a constant $K_{\text{Mem}}$ as $M(n) = K_{\text{Mem}} *n.$

In (2.21a), the first input word The is associated with the memory cost of $M(0)$ MUs for the prediction of the matrix subject. The memory cost incurred at the processing of the next input reporter is 0 MUs, relating to the prediction of the matrix verb, which is always assumed to be zero. At the next parse state, $2M(0)$ MUs in total is associated with the word who, corresponding to the prediction of the embedded verb inducing $M(0)$ MUs and the prediction of an empty category NP coindexed with the wh-pronoun who resulting in another $M(0)$ MUs. The following word the is associated with the memory cost of $3M(0)$ MUs, since three predictions have been made: the same two predictions as at the previous state, plus one additional prediction of the embedded subject noun causing extra $M(0)$ EUs. The memory cost associated with the next parse state, at the word senator, is $2M(1)$ EUs, where the two predictions, one for the embedded verb and the other for the NP-gap, are still not satisfied. At this stage, there is one new discourse referent “the senator” which has been processed since the head noun senator was initially predicted at the parse of the previous word the, therefore
the memory cost for each of the two predictions increases to \( M(1) \) MUs, resulting in the total memory cost of \( 2M(1) \) MUs for the parse. For the next word **attacked**, 0 MUs is incurred because the embedded verb and empty NP predictions are now satisfied and only the prediction of the matrix verb still remains, which however costs 0 MUs. After processing the next input **admitted**, the matrix verb prediction is satisfied, and a prediction of an object NP is made but no new discourse referent intervenes, inducing a memory cost of \( M(0) \) MUs. At the following word **the**, still the object NP prediction is not satisfied, leading to \( M(0) \) MUs again. At the final parse state, the input **error** is attached into the current structure, satisfying the object NP prediction and no prediction is made, in consequence 0 MUs is caused.

As can be seen from (2.21b), the memory cost profile for the subject-extraction is identical to that for the object extraction apart from the region between the embedded verb **attacked** and its object noun **senator**. The memory cost associated with the word **attacked** is \( M(0) \) MUs, corresponding to the prediction of an object NP. The memory cost associated with the next parse state, at **the**, is \( M(0) \) MUs too, matching to the same prediction as at the previous state. At the next parse state, the object NP prediction is satisfied when the input word **attacked** has been processed, resulting in a 0 MU memory cost at this point.

The memory cost for (2.19) predicts that processing the object-extracted relative clause construction (2.19a) requires more memory resources than its subject-extracted relative clause version (2.19b), since according to (2.21), the maximal memory complexity for (2.19a) is \( 2K_{mem} \) MUs at the point of processing the embedded NP subject **senator**, whereas the memory cost for (2.19b) is, in contrast, zero throughout all the parse stages.\(^9\)

\(^9\)In his paper, Gibson (1998: 34) measures the maximal memory complexity of the subject-extraction (2.19b) as \( 2M(0) \) MUs, which occurs at the point of processing the relative pronoun **who**. This seems to be a contradiction since \( 2M(0) \) MUs must be zero. This assumption is an oversimplification, as Gibson admits (1998:28-30), because a more accurate indication of memory complexity is given by the sigmoidal version of the memory cost function, according to which \( M(0) \) MUs should have some positive value. In this sense Gibson's interpretation that the maximal memory complexity of the subject-extraction (2.19b) is at \( 2M(0) \) MUs is proper.
2.4.1 Discussion

(2.22a) Integration and Memory cost for (2.19a)

<table>
<thead>
<tr>
<th></th>
<th>The reporter</th>
<th>who</th>
<th>the</th>
<th>senator</th>
<th>attacked</th>
<th>admitted</th>
<th>the</th>
<th>error.</th>
</tr>
</thead>
<tbody>
<tr>
<td>EUs</td>
<td>-</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>MUs</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Maximal complexity: 3*K\textsubscript{Int} EUs at attacked and admitted; 2*K\textsubscript{Mem} MUs at senator

Total complexity: 7*K\textsubscript{Int} EUs + 2*K\textsubscript{Mem} MUs

(2.22b) Integration and Memory cost for (2.19b)

<table>
<thead>
<tr>
<th></th>
<th>The reporter</th>
<th>who</th>
<th>attacked</th>
<th>the</th>
<th>senator</th>
<th>admitted</th>
<th>the</th>
<th>error.</th>
</tr>
</thead>
<tbody>
<tr>
<td>EUs</td>
<td>-</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>MUs</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Maximal complexity: 3*K\textsubscript{Int} EUs at admitted; none for Memory cost

Total complexity: 6*K\textsubscript{Int} EUs

As can be seen from (2.22) above, the combined version of the integration and memory complexity profiles, the SPLT integration and memory theory correctly predicts the higher complexity of the object-extracted relative clause (2.19a) as compared with a subject-extracted relative clause counterpart (2.19b). This is evidenced by the difference in the maximal complexity between the two sentences; the maximal complexity for memory cost of (2.19a), measured at 2*K\textsubscript{Mem} MUs, is greater than that of (2.19b), namely zero, yet the maximal complexity for integration of both the sentences is equal at 3*K\textsubscript{Int} EUs.

One advantage of the complexity measuring system assumed by the SPLT is that it is able to compute the maximal quantity of memory resources that are required at any point during the parse of a sentence. Hence, the theory should be well-suited to account for a processing overload effect as it can pinpoint the parse stage at which processing breakdown occurs.
Nevertheless, the relationship between the two components of the SPLT, integration cost and memory cost seems unclear, although these concepts attempt to reflect the relationship between the available memory resources of working memory and the sentence processing mechanism. This is because although the SPLT hypothesises that linguistic integration and storage access the same pool of resources, the theory has not yet demonstrated whether an energy unit (EU) could be converted into memory unit (MU) without applying time unit (TU), and if so, then how this could be mathematically formulated. This makes it impossible to determine which is the main cause for syntactic difficulty - linguistic storage or computation - and also to determine the absolute maximal complexity point of processing because there are two potential points of maximal complexity, one for integration and the other for memory. For instance, as shown in (2.22a), the maximal integration complexity for processing (2.19a) is $3K_{\text{int}}$ EUs at the stages of parsing attacked and admitted, and the maximal memory complexity for (2.19a) is $2K_{\text{stem}}$ MUs that occurs at senator. Without a formula for converting energy units into memory units, it is difficult to recognise which component of the theory causes relative syntactic difficulty of the object-extraction construction (2.19a) and also at which point of the parse the memory resources are most expended for processing.

Finally, in relation to the formula for the timing of linguistic integration there seem to be two main factors causing incomplete syntactic processing: (i) $M_{\text{current-memory-used}} > M_{\text{capacity}}$ = a working memory overflow, whereby the memory resources used for storage exceed the whole working memory capacity of the listener. (ii) $I_{\text{struct-integ}} > M_{\text{capacity}} - M_{\text{current-memory-used}}$ = inexecutable linguistic integration, where the quantity of memory resources required to perform the integration exceeds the working memory resources available for integration. Under the circumstance of (ii), any integration is theoretically possible, but integration time required is too long to operate processing, so that the listener would not be able to hold attention. However, (ii) necessitates a conversion formula with which energy resources (in EUs) can be converted into memory resources (in MUs) without employing a time unit (TU) as designated in the relationship between memory and integration cost. It would be an interesting topic to identify which component causes processing breakdown for each processing difficulty phenomenon and the average maximum time over which syntactic processing is unable to proceed due to the lack of attention.
2.5 A summary of the views of short-term memory

Short-term or working memory is one of the most researched topics in the field of cognitive psychology and neuroscience since the sixties. Although there are diverse models and theories of working memory in abundance, in this section we shall briefly consider two specific models of working memory which are especially popular among psychologists, namely those of Gathercole & Baddeley (1993) and Caplan & Waters (1999), and assess the importance of working memory as a parser that linguists would be interested in.

Fundamental ideas of two types of storage in human memory, short-term memory and long-term memory, were first described in the late fifties (Broadbent 1957, 1958), and were crystallised into a definite theory by Atkinson and Shiffrin (1968). Although their work is now of only historic interest, its influence can still be seen in many current theories such as SAM developed by Shiffrin (Gillund & Shiffrin 1984).

Short-term memory was conceived to be a temporary storage to retain a small amount of information, e.g. a telephone number of seven digits. The procedure to retain information in short-term memory is to rehearse, i.e. to repeat the information to oneself over and over again. Long-term memory, by contrast, was considered as a high-capacity permanent storeroom, into which information held in short-term memory could be transferred by rehearsal. Since rehearsing information each time provides a chance for it to get into long-term memory, increased rehearsal of the information insures increased likelihood of retention in long-term memory. Also, taking a piece of new information into short-term memory with a limited capacity entails the displacement or loss of old information. (2.23) based on Anderson (1995: 28) illustrates the relationship between short-term and long-term memory.

(2.23)
The theory was examined particularly in one of the experimental paradigms called 'free recall', whereby a list of words are read out to subjects, and they start to recall the words in free order. The results of this type of experiment produce the so-called serial position curve that shows the probability of recall as a function of the position of the word in the list. The characteristic of this curve is bell-shaped and the recall accuracy at the beginning (primacy effect) and at the end (recency effect) are better than that at the mid region.

Atkinson and Shiffrin's theory justifies the shape of this serial position curve; the worse performance in the middle is due to the supposed deletion of a word already held in short-term buffer in order to receive another new word; the primacy effect should arise because the first few words do not have to compete with previous words, so they tend to be more rehearsed; the recency effects can also be explained by the fact that the last few words are more likely to stay in the short-term buffer, and receive better recall.

Although the Atkinson and Shiffrin theory seems to fit such simple experimental paradigms as free recall, more complex experiments have given evidence against the theory; passive rehearsal of material will not increase its recall (Craik & Watkins 1973; Glenberg, Smith & Green 1977, Neisser 1982), but deeper processing does (Craik & Lockhart 1972; Craik & Tulving 1975); both acoustic and semantic information can influence memory performance (Bower & Springston 1970; Potter & Lombardi 1990).

An alternative theory to Atkinson and Shiffrin's two-memory assumption is that one general memory system processes information obtained by sensory systems (Melton 1963; Wickelgren 1974; Ericsson & Kintsch 1995); the general consensus seems to be that long-term knowledge plays an integral role in working memory performance in that the contents of working memory consists primarily of currently activated long-term memory representations and can also extend to those that can be quickly reactivated. (Miyake & Shah 1999). Concerning this subject, I shall briefly explain below the concept of working memory proposed by Baddeley (Baddeley & Hitch 1974, Baddeley, Thomson & Buchanan 1975, Baddeley & Lewis 1981, Baddeley 1986, Gathercole & Baddeley 1993).

Baddeley and Hitch (1974) used the term, working memory, to describe the
short-term memory that is involved in the control and maintenance of temporary information. They suggested that working memory plays a central role in the service of complex cognitive activities such as language comprehension. Regarding the relation of working memory to long-term memory, Baddeley & Logie (1999) assume that these two memory systems are functionally separable in that a major role of working memory is to retrieve stored long-term knowledge relevant to the tasks in hand and to encode the outcome of its operations into long-term memory.

The working model presented by Gathercole & Baddeley (1993) has a tripartite structure consisting of a central executive and two slave systems, the phonological loop and the visuospatial sketch pad, so this model has an inherently non-unitary nature. A representation of the model in question is given below in (2.24), which is based on Gathercole & Baddeley (1993: 4).

![Diagram of the working memory model](image)

The central executive component offers the mechanism for control processes in working memory, fulfilling many different functions including recall of length lists of digits (Baddeley & Hitch 1974) and logical reasoning (Baddeley & Hitch 1974). In terms of language comprehension, Gathercole & Baddeley (1993) also assume that the central executive plays an important role of activating representations in long-term memory extending up from individual words and concepts to complex schemas, processing syntactic and semantic information and storing products of processing.

The regulatory function of the central executive is thought to be guided by a system for the control of attention; familiar activities are guided by schemas which are triggered by environmental cues, but the Supervisory Attentional System takes over in order to override the routine process by inhibiting or activating schemas directly. The idea of control of attention seems to be closely linked to Cowan's model of working memory (Cowan 1988, 1995), which is more unitary than Baddeley's in the sense that Cowan's concept of memory subsumes sensory and abstract activation based on any modality in
any form of representation. (Thus, his working memory involves all information relevant to a task, including elements inside the focus of attention, which is included in working memory, those out of the focus nonetheless temporarily activated in long-term memory, and inactive elements of memory with activated retrieval cues.)

The phonological loop is a specialised temporary memory system used to maintain verbal memory traces, which decay with time, by a process of subvocal rehearsal. It should be noted that this rehearsal system and short-term memory are different in that the rehearsal system is not a halfway station to a memory network, and information does not go through the rehearsal systems to activate a permanent concept in the permanent memory network.

In terms of language comprehension, views on the function fulfilled by the phonological loop varies, although most current theories agree on the assumption that syntactic and semantic processes proceed on-line without reference to a phonological working memory representation, so they are handled by the central executive, as stated above. One view of the phonological loop is that it provides a backup representation for incoming linguistic input that can be consulted during off-line linguistic analysis when the input is long or syntactically complex such as passive and centre-embedded sentences (Baddeley, Vallar & Wilson 1987) or new and needing to be learned (Baddeley, Gathercole & Papagno 1998).

An alternative view argued by Waters et al. (1987) is that phonological memory presentations do not provide an input into the initial syntactic and semantic analysis, but into the subsequent checking of the parsed syntactic structures. On the basis of considerable evidence for a division of the central executive into verbal and visual components (Shah & Miyake 1996), Caplan & Waters (1999) argue for the possibility that the verbal working memory constitutes a separate subsystem within it specialised for syntactic processing, i.e. assigning syntactic structure and using it to determine the meaning of a sentence.

Their studies in Caplan & Waters (1999) showed that concurrent verbal memory loads do not disproportionately affect comprehension of syntactically more complex sentences, either in normal subjects or in subjects
with extremely reduced working memory capacity or in aphasic patients. Based on these results, they suggested that there is a specialisation in the verbal working memory system involved in syntactic processing that is separate from that measured by standard tests of working memory.

Syntactic processing is one stage of the entire language comprehension process: some of the other kinds of operations in the process are acoustic-phonetic conversion, lexical access, intonational contour recognition, semantic and pragmatic processing, which normally act in a cooperative way (Marslen-Wilson 1987). Considering that the language comprehension process is integrative in nature and is highly practised in human cognitive functions, Caplan & Waters hypothesised that "one resource system is utilised by all these different types of processes that combine in the interpretation process" (Caplan & Waters 1996b; Waters & Caplan 1996).

In summary, there seems to be a general consensus amongst psychologists about the role of working memory in language comprehension. Baddeley and colleagues assume the working memory model comprising three components: the central executive, the phonological loop and the visuo-spatial scratch pad: for language processing, the central executive allocates attention to a comprehension task, performs computational functions and stores information, while the phonological loop, a slave system of the central executive, stores verbal information when the central executive is overloaded. In contrast, Caplan & Waters assume that a finer division in the verbal working memory, which is a division of the central executive of working memory, is responsible for syntactic and semantic processing. (They also argue that working memory contains specialisations for different verbal processes.)

However, the issue of the specialisation for syntactic processing in the working memory system seems of no concern to linguists, so long as it is agreed that syntactic processing takes place inside working memory. The more important question relevant to linguists, or rather to this thesis, is about the limits of working memory capacity: Are there limitations of a working memory resource that is utilised by syntactic processing? Baddeley and colleagues assume that there is limited capacity for activation, rehearsal and so forth, whereas Caplan & Waters hypothesise a single working memory resource for the entire set of operations for language processing, as
mentioned above. My thesis is based on the assumption that working memory has a limited resource available for syntactic processing, which seems uncontroversial among psychologists, and explores what kinds of syntactic factors are responsible for consuming the resource.
3 A Word Grammar parser

In the previous section, we have seen some theories of parsing, all of which have assumed the grammar and lexical representations based on phrase-structure grammar, e.g. X Theory (Chomsky 1986b). Hence, syntactic complexity and processing load are associated with the number of phrasal nodes.

In this section, I shall introduce an alternative parsing theory based on a syntactic theory called Word Grammar (Hudson 1984, 1990, 1992, 1994a, 1994b, 1996a, 1996b, 1997b). This is a particular version of dependency theory, whereby a sentence is structured in a monostratal fashion without any phrases at all. Firstly, the assumptions concerning sentence-structure made by Word Grammar are given. This is followed by an introduction to the theory of parsing which represents such syntactic structures. Thirdly, a method of measuring syntactic complexity, namely 'dependency distance' is explained. Finally, I will attempt to answer two questions: (i) What, if any, are the fundamental differences between a parser assuming phrase-structure grammar and a Word Grammar-based parser? (ii) Which model of sentence-structure should be preferred from the parsing viewpoint?

3.1 Word Grammar

The theory sees an individual word as the minimal and also the maximal syntactic unit of a sentence which enjoys pairwise relationships with other single words, excepting the treatments of coordination, compounding and clitics. A sentence having \( n \) words therefore is structured as a series of \( n-1 \) pairwise combinations of words. These unequal relations are termed dependencies, where one word of a pair is the 'parent' and the other word is the 'dependent'. The types of dependencies are sorted according to their syntactic functions such as 'subject', 'object', and 'raiser'. The dependency relations of (3.1) are exhibited below.
The sentence structure of (3.1) is represented by the individual words tied to other words by the dependency arrows. The dependencies must be linked in such a way that they are compatible with the information on individual words in the grammar. In the following sections, we shall see how the grammatical information is codified in the Word Grammar system.

3.1.1 Default Inheritance

In Word Grammar a grammar is a network of concepts for words and other entities which interrelate with one another in diverse ways. These concepts are generalised and arranged in hierarchies by logical operations ascribed to Default Inheritance. This mechanism, which is commonly used in Artificial Intelligence and Computer Science, associates two concepts together: a model (generalisation) and its instance, in which if a concept A is a model of a concept B, then B is an instance of A. Inheritance allows the instance to inherit all the properties of the model by default. Inheritance hierarchies, therefore retain the most general concepts at the top and the most specific ones at the bottom. The inheritance principle is formally defined as follows:

\[ \text{Inheritance} \]

If B is an instance of A, then B affiliates itself to any valid proposition P which also applies to A with the substitution of B in P for A.

The hierarchical relation of Inheritance can be expressed by the predicate 'isa': iff B isa A, then A is a model of B. As observed in (3.2), the super-category A is on top of the base of the triangle. In contrast the sub-category B is beneath the lower end of the vertical line whose upper end comes in contact with the triangle's apex to the bottom. It denotes that this simple diagram composed of a triangle with a straight line joined at its apex allows A's
properties to be inherited by B.

(3.2)

\[ A \downarrow \]
\[ B \]

A isa hierarchy for 'B isa A'

Central inheritance hierarchies in Word Grammar are the organisation of word-classes comprising lexemes and the organisation of dependency types. Inheritance through the isa hierarchy is one way to deduce new information from old, which is highly relevant to syntactic processing as its first step. It shall be elucidated later.

(3.3)

\[ \text{word} \]
\[ \text{noun} \quad \text{preposition} \quad \text{verb} \quad \text{adjective} \quad \text{adverb} \quad \text{coordinator} \]
\[ \text{pro} \quad \text{common} \quad \text{proper} \quad \text{full} \quad \text{auxiliary} \quad \text{gerund} \]
\[ \text{noun} \quad \text{noun} \quad \text{noun} \quad \text{verb} \quad \text{verb} \]

A partial hierarchy of the Word Grammar word-classes for English

(3.4)

\[ \text{dependent} \]
\[ \text{pre-dependent} \quad \text{post-dependent} \quad \text{valent} \quad \text{adjunct} \]
\[ \text{extractee} \quad \text{subject} \quad \text{extraposee} \quad \text{complement} \quad \text{pre-adjunct} \quad \text{post-adjunct} \]
\[ \text{sharer} \]

A partial hierarchy of the Word Grammar dependency types for English
3.1.2 Overriding

In the case that an instance has a more specific proposition which is contradictory to a property of its model, it is required for the model to disinherit it. To accommodate exceptions in the default inheritance system, current Word Grammar utilises 'automatic overriding'. A proposition which details some specific property overrides any more general one which would otherwise be inherited. Here is a simple example of default inheritance and overriding of defaults by automatic overriding.

Inheritance: espresso inherits [1] and [2].

[1] Coffee is dark brown.

Overriding defaults: decaffeinated white coffee does not inherit [1] and [2].

[5] Decaffeinated white coffee is coffee.
[6] Decaffeinated white coffee is white-ish.
[8] Decaffeinated white coffee is dark brown and contains caffeine.

There used to be 'stipulated overriding' which was supported by Word Grammarians (Hudson 1990, Fraser and Hudson 1992) whereby a simpler treatment for alternatives was possible. This strategy employs a special predicate NOT that blocks inheritance of general properties. Handling an irregular past-tense of a verb like SHOOT we need to stipulate not only the following proposition (3.5) but also (3.6).

(3.5) NOT: structure of past SHOOT = stem of SHOOT + mEd.

(3.6) structure of past SHOOT = <shot>.

On the other hand, stipulated overriding offers an elegant solution to alternatives, with which automatic overriding may allegedly have a severe problem. Take LEAN as an example. There are two alternative forms for its past tense, the regular form leaned and the irregular one leant. To accommodate these two forms, stipulated overriding required only one
proposition (3.7).

(3.7) structure of past LEAN = <deant>

Nevertheless there are reasons why the first system, automatic overriding, is presently in favour in Word Grammar. First, this operation is a standard one upheld by the majority of linguists. Secondly, it is generally true that more specific information does block inheriting any more universal proposition, which means that the number of problematic cases like alternatives are limited. Thirdly there is in theory a way to handle alternatives without stipulated overriding in Word Grammar (although this might not be appealing), namely by recognising distinct sub-types. It appears plausible on the grounds that different morphology occasionally attracts divergent meanings. HANG is one of the examples. Its two alternative past-tense forms, namely *hanged* and *hung* are disparate in meaning - 'execute' and 'attach in a high position'. The diagram (3.8) below demonstrates the treatment for the alternatives of HANG by automatic overriding.
The fourth point in favour of automatic overriding is that it is easy to implement in an algorithm for processing. Guaranteeing that the processor always accesses more specific information before more general entails that the first solution the processor gains is always the right one. This is because processing is always bottom-up in terms of the isa hierarchy and given "monotonic" processing, it builds structures up without changing or destroying them. The flowchart modelling the recognition of *hung* and *hanged* is displayed below.
stage 0: perception of an input

stage 1: recognising it as a word

stage 2: searching for the model of Wx

stage 3: matching Wx with its model by the Best Fit Principle

stage 4: inheriting the properties of the model

HANG/1:past → Wx

stage 0: perception of an input

stage 1: recognising it as a word

stage 2: searching for the model of Wx

stage 3: matching Wx with its model by the Best Fit Principle

(morphological analysis of "hanged")

stage 4: inheriting the properties of the model

HANG → Wx
3.1.3 Temporary words and permanent words

So far we have seen how a particular concept inherits its generalities from its model concept through inheritance hierarchies in the network of a grammar. Now one may come to the question of how syntactic parsing relates to the default inheritance operation. Suppose you hear a string of sound. If this pattern of sound was an utterance consisting of some words of your native tongue, your brain would mandatorily proceed on phonological processing of the utterance, which would translate the acoustic input into a string of postulated words. Then at the next level you would start to parse the word string, that is, to establish a syntactic structure in it.

The first step of parsing by a Word Grammar parser should be to incorporate each word of a string of words into the inheritance hierarchies. Suppose you perceive the utterance 'meadows behind green trees'. Each word of the utterance is recognised at a different point over a stretch of time and is dealt with incrementally. This means that these four words under process are not stored in your permanent memory but are temporary ones. They are associated with temporary concepts signifying particular things which facilitate your understanding what the utterance means. I shall provide names for these temporary words and concepts following the convention of employing the prefixes 'w' and 'c' respectively followed by numbers. Hence in the utterance above there are four words 'w1' to 'w4' to be processed which have as referents the distinct concepts labelled 'c1' to 'c4' respectively.

In order for temporary words to integrate into the inheritance hierarchies, they need to be recognised as instances of their corresponding permanent words considering that the 'isa' relationship is the basis of all hierarchical classification structures. These stored words hanging at the bottom of the hierarchies are christened lexemes or sublexemes for singular, plural, or irregular forms, if the lexemes have any. They are notated in capital letters throughout. A 'dictionary word' lexeme may contain more than one word-forms on the grounds that they share the same permanent concept in a grammar network. For instance, MEADOW deals with both the singular and the plural forms, namely meadow and meadows, while a lexeme for an adjective or an adverb etc. possesses a single word-form in English.
The figure (3.9) below displays the temporary words, their lexemes and their temporary concepts for the utterance 'meadows behind green trees'. Notice that a word in single quotation marks in the figure indicates the general concept, the sense of a referent. The token words prefixed by 'w' are needed not only for processing but also because they may have varied pronunciations, word-shapes, grammatical relations to other words even if they share the same lexical item. Take 'green recruits looked green.' as an example. 'w1' and 'w4' of the phrase have the identical word-form green. Despite belonging to the lexeme GREEN, they must be distinguished on account of their having different grammatical functions in the phrase, viz. 'w1' is a pre-adjunct of 'w2' and 'w4' is a complement of 'w3'.

(3.9)

**meanings**

- 'meadow'
- 'behind'
- 'green'
- 'tree'

words

- MEADOW: singular
- MEADOW: plural
- BEHIND
- GREEN
- TREE: plural
- TREE: singular

utterances

- meadows
- behind
- green
- trees

The temporary words, their lexemes, referents, and senses for meadows behind green trees

3.1.4 The Best Fit Principle

As mentioned before, unification of temporary words with the end-concepts of the isa hierarchies, namely, lexemes is the initial step of syntactic parsing. When you hear a sound pattern of speech or see a string of letters, your brain will automatically and extremely fast attach it to the lexeme whose prototypical sound or word shape is most similar. This binding operation as
a part of understanding is a mystery in the field of psychology. It is however, feasible to state a principle which governs the procedure of horizontally linking a linguistic input with some concept in the grammar. The principle arises from the more general one defined as the Best Fit Principle by Hudson.

**The Best Fit Principle**

An experience E is interpreted as an instance of some concept C if more information can be inherited about E from C than from any alternative to C. (Hudson, 1990: 47)

Thus as soon as word tokens are joined to lexemes according to the Best Fit Principle, cf. the emboldened links in (3.10), the default inheritance mechanism allows the temporary words to load attainable properties from the stored words such as their word-types and dependency types. Using the previous example (3.1), I present below the steps in parsing by linking operations of default inheritance supplemented with the Best Fit Principle.

(3.10)

```
words

|| w1 | w2 | w3 | w4 | w5 |
---|---|---|---|---|---|
      | WALK:pres | ON | GILDED | SPLINTER:plural |
      | w1 | w2 | w3 | w4 | w5 |

utterances

|| walk | on | gilded | splinters |
---|---|---|---|---|
      |       |   |     |          |
```

Unification of the temporary words and permanent words for *I walk on gilded splinters*.

a. default inheritance in word-class

1. w1 isa l.
2. l isa pronoun.
3. :: w1 isa pronoun.
4. w2 isa WALK:pres.
5. WALK:pres isa present-tensed verb.
6. :: w2 isa present-tensed verb.
7. w3 isa ON.
8. ON isa preposition.
9. w4 isa GILDED.
10. w4 isa adjective.
11. :: w4 isa adjective.
12. w5 isa SPLINTER:plural.
13. :: w5 isa SPLINTER:plural.
14. SPLINTER:plural isa SPLINTER.
15. SPLINTER isa common noun.
16. :: w5 isa common noun.
[9]  \( \therefore \) w3 isa preposition.

b. default inheritance in parent-demand

[1]  w1 isa l.
[18]  A noun isa word.
[19]  A word has a parent.
[21]  \( \therefore \) w1 has a parent.

[22]  A present-tensed verb isa tensed verb.
[23]  A tensed verb is a finite verb.
[24]  A finite verb is a verb.
[26]  A finite verb may have a parent.
[27]  \( \therefore \) w2 has a parent.

[7]  w3 isa ON.
[8]  ON isa preposition.
[29]  A preposition isa word.
[19]  A word has a parent.
[30]  \( \therefore \) w3 has a parent.

[10]  w4 isa GILDED.
[31]  An adjective isa word.
[19]  A word has a parent.
[32]  \( \therefore \) w4 has a parent.

[14]  SPLINTER:plural isa SPLINTER.
[19]  A word has a parent.
[34]  \( \therefore \) w5 has a parent.

c. default inheritance in dependent-demand

[22]  A present-tensed verb isa tensed verb.
[35]  A tensed verb has a subject.
[36]  \( \therefore \) w2 has a subject.

[7]  w3 isa ON.
[8]  ON isa preposition.
[37]  A preposition has a complement.
[38]  \( \therefore \) w3 has a complement.
d. default inheritance in word-order

[40] A word's dependent follows it. [39] A post-dependent is a dependent.
[41] A pre-dependent is a dependent. [40] A word's dependent follows it.
[43] A word's pre-dependent follows it. [38] w3 has a complement.
[51] w3's complement follows it.
A subject is a pre-dependent.
A word's subject precedes it.
A word's subject follows it.
w2 has a subject.
w2's subject precedes it.
w2's subject follows it.

(3.11) is a schematic representation of what is inherited by default inheritance, i.e. full word-entries, via temporary words on top of the sentence under process.
3.1.5 Word Grammar universal principles

If we compared (3.10), the complete dependency structure of I walk on gilded splinters, to (3.11) where the information about each word is supplied solely by default inheritance, the most significant difference between the two is that none of the latter’s dependency arcs are linked. Although the illustration of (3.11) is inaccurate owing to disregarding the incrementality of parsing, it does at least represent the same stage of parsing for each word, namely the completion stage of default inheritance. The next stage is to provide each word being processed with a coherent dependency structure compatible with the grammar. In Word Grammar, general principles (Hudson 1996a), which are assumed to be universal, govern sentence-structures consisting of words linked in a pair-wise fashion. Making use of these principles, the parser executes linking and selects the most plausible dependency structure out of the enormous number of possibilities.

**The No-dangling Principle**

Every word must depend on a parent.

This principle may not be essential for the parser because each word automatically inherits the property of having a parent. This point is brought home by a vertical arrow with a dot on top of it in (3.11). (The dot stands for an unidentified parent.) As soon as the parser parses a word, its vertical line is linked with its parent.

**The Sentence-root Principle**

Every sentence has only one sentential root unless it is coordinated with another.

The parent of a word that is a sentence-root is not an actual word but a potential one. In (3.11) walk is the sentence-root, therefore it is held by a plain vertical arrow, whose length is assumed to be infinite. If several sentence-roots are coordinated within a sentence, the number of sentence-roots is equal to that of the coordinated participants as shown in the example (3.12) below.
This principle combined with the no-dangling principle provides the parser with two vital constraints. The first bans any word from being detached from its parent; this is assured by the no-dangling principle as well as default inheritance. Secondly it prevents any word that is inside one parent-dependent relationship from being a parent to or dependent of a word outside this relationship. This restriction arising from the no-tangling principle should also be seen in the context of the sentence-root principle. Because a sentence-root arrow is presumed to be stretched infinitely to the perpendicular, any dependency arrow passing across a sentence-root is illicit. A sentence-root arrow divides the dependency structure of a sentence into two territories, namely the one before the sentence-root and the other after it. This contributes to the great ease of the parser’s task, since once the parser hits a word that is a sentence-root, parsing after it will be done without access to any words in the pre-sentence-root area. Coordinations and surface dependency structures are dealt with later.

The Competition Principle
In surface structure, each word depends on no more than one word.

The recent framework of Word Grammar recognises two types of dependencies: surface dependencies and extra dependencies. The former are responsible for word order of a sentence, hence any link belonging to this type is tangle-free thanks to the no-tangling principle. All the surface dependencies in a sentence constitute its surface dependency structure. Any other dependencies which are unable to be accommodated in a surface
structure belong to the latter type. As the names of these dependencies suggest, surface dependency arrows appear on the surface of words (i.e. above them) while extra dependency ones appear under them. The figure (3.13) below illustrates this principle.

To parse the sentence *It will rain*, each word is initially supplied with dependency details by the isa hierarchy as shown above. When the word *will* is input, there are two tasks for the parser to carry out. Firstly, it unifies the parent of *It* with the subject of *will*. Secondly, the subject arrow projecting from the sharer is demoted from the surface dependency structure as it has lost the competition for a place in the surface structure. This is not only because the subject relation in the surface structure has already been established for *it*, but most crucially because the subject arrow in question would be tangling with the sentence-root arrow, which is against the no-tangling principle. Hence this subject arrow appears below the words and shares the subject *It* with the subject arrow in the surface structure. The state of the parse after the second word *will* is presented in the figure (3.14) below.
The final word of the example sentence *rain* is then input. It is first identified as the sharer of *will*. The subject arrow extending from the input, however cannot be accommodated in the surface structure for the reasons stated above. Consequently the subject arrow is suppressed in the extra dependency structure and linked with the subject. The parse of the sentence is successfully completed, and the diagram is displayed below.

\[(3.15)\]

The **Raising Principle**

If a word \(w_d\) depends on \(w_p\) as well as a sharer \(w_r\), and \(w_r\) depends on \(w_p\), then \(w_d\)'s surface parent must be \(w_p\).

This constraint controls the choice between surface and extra dependencies when sharers occur in a surface structure. The parser might operate without exercising this principle, because any dependency analysis in defiance of this principle seems to result in ungrammaticality due to the violation of other principles like the no tangling one. This means that the parser can be designed in such a way that it allocates the earliest parent to the surface structure at least for English. The parser would always attain the right linking by adopting this strategy. The figure (3.16) below shows an illegal dependency analysis, which infringes the raising principle, but more fundamentally defective information has been inherited: the sentence-root is not on *will* but on *rain*. 
Even for noticeably exceptional constructions like VP fronting, on which there is a remark in Hudson (1996a, p.p.24-25), the parse is efficiently executed without the raising principle. Take the sentence Roll on the floor you must as an instance. After the input string Roll on the floor you has been parsed, the last item in the input cannot be identified as the subject of Roll in the surface structure in as much as subject is pre-dependent; the subject's position is fixed in surface structure. Thus the subject arrow is redirected to the lower structure in anticipation of being supported by a subject dependency in the surface structure. At the point of parsing the auxiliary verb must, Roll and you are linked with this current input as its sharer as well as extractee and subject respectively in the surface structure. The subject link in the extra structure is subsequently attached to you by virtue of Roll being the sharer of must. The processing stages for the last three words of the example sentence are exhibited in the illustration (3.17) below.
The Surface-structure Principle

Word-order constraints including the no-dangling, no-tangling, and sentence-root principles are only valid within surface structure.

Having established two kinds of dependencies: surface and extra ones, we need to formally state which principles have effect on which dependencies. These principles demand continuous phrases exclusively in surface structure. Surface dependencies are therefore roughly equivalent to trees of Phrase-structure grammar in the sense that discontinuous sequences are not allowed in either.

As we have seen, the parser automatically initially builds extra dependencies always in the surface structure. Three general types of extra dependency
motivated by discontinuity are considered in the following to examine whether constructing extra dependencies is costly in terms of the parser's memory capacity.

The first case is raising, which we have already encountered in the preceding examples. Its characteristic complement termed sharer shares another valent of the raising verb. This doubling-up dependency complies with information on valency of the raising verb in the lexicon, and the parser will mechanically be able to categorise this dependency as an extra one. The illustrations (3.18) and (3.19) below present two patterns of a sharer: the one sharing subject of the raising verb and the other sharing its object.

(3.18) (3.19)

sharer's subject doubling up as
subject of tend

sharer's subject doubling up as
object of expect

The second construction is extraction where a front-shifted word having a valent relationship with the verb is shifted from its normal position to the front. Take the following sentence for example: Who do you think you are talking to? At the point of parsing talking, the extractee, Who, has still not been identified as a legitimate dependent. In consequence, every time a verb enters as input, in this case do, think, are, talking, the extractee
relationship is recursively passed down the surface dependency chain until the extractee can be associated with a word as its valent in the surface structure. So in processing the next word to in the example, the parser recognises Who as its complement, and establishes the complement plus extractee links in the extra dependency structure. The extractee dependencies are connected between Who and any parent from which the extractee relation is passed down until the extractee is, as it were, cashed as a rightful dependent. Hence the parser's operation involving extraction is creating extra dependencies from an extractee at each point where the extractee has a new parent until they form an extraction chain, which would be computed with little cost. The diagram (3.20) below provides the dependency structure for the example sentence.

![Dependency diagram](image)

Thirdly, as for extraposition from NP, more extensive search seems to be needed in order to discover an exact extra dependency, because there are more possibilities to choose from. Yet the potential antecedent of an extraposee is normally the nearest noun dependent of a verb which is also a parent to the extraposee. Thus the parser would only have to check the nearest nouns which share the same parent as the extraposee. The extra dependencies induced by extraposition can be almost automatically construed from the surface dependency structure. The figure (3.21) below shows the analysis of a sentence which involves extraposition.
As we have observed above, the extra dependencies associated with discontinuous construction, e.g. raising, extraction, extraposition, are effortlessly reconstructed by the parser thanks to their predictability. A parent of a sharer is also a parent to some surface dependent which is dependent on the sharer in extra structure. For extraction, an extractee's parent is an ancestor of a word that is on the other hand a parent to the extractee in extra structure. Regarding extraposition constructions, a very local search is called for to identify a nominal element as the ancestor of an extraposee out of its parent's neighbourhood. This rather straightforward procedure for extra dependency-making forces us to surmise that extra dependencies are build up virtually at no cost. I therefore assume that it is not important to know precisely what extra dependencies add to sentence processing load.

**THE COORDINATION PRINCIPLE**

Inside a coordination, the no-tangling principle applies to each conjunct separately.

This principle helps the parser to handle coordinations. Suppose the following sentence is being processed: *She finished her Perrier and sucked the lemon.* At the point the coordinator *and* is parsed, it serves to signal that there is a coordination in the sentence. Then the next word *sucked* is input, the parser joins this word and *finished* as conjuncts by virtue of having the same word class. Since *sucked* still needs a subject, and cannot have one exclusively belonging to it, it ought to share the same subject as *finished*. Linking *She* with *sucked* de facto involves tangling with the sentence-root arrow projecting from *finished*, yet this is sanctioned by the coordination principle because this tangling involves a different conjunct.
The diagram (3.22) below describes the dependency structure for the example sentence. Notice that \textit{and} is merely an indicator of a coordination so that it does not depend on any word.

\begin{equation}
\text{(3.22)}
\end{equation}

This functional explanation for Word Grammar's universal principles in terms of their effects on parsing concludes with a brief summary. The primary objective of a Word Grammar parser is to reconstruct a surface structure for an input string of words, on the assumption that extra dependencies cost almost nothing. In the course of a parse, the parser should abide by the following rules which are regulated by the universal principles:

1. Every word in a sentence must be a dependent of one and only one parent with the exceptions of sentence-roots and coordinators. The former depend on virtual words and the latter have no parents.
2. In surface structure, dependencies must not tangle except where tangling is due to coordination.
3. If \(w_z\) depends on \(w_x\) as well as \(w_y\) and at the same time \(w_y\) depends on \(w_x\) then the dependency between \(w_z\) and \(w_x\) takes precedence over the one from \(w_z\) to \(w_y\) in surface structure.

### 3.2 Word-Grammar based parsing

In this section, we shall examine one parse of a sentence step by step to get a clear picture of how the Word Grammar parser processes sentences. First of all, we shall review the basic procedure of parsing as a preliminary. The parser works incrementally from left to right (e.g. Marslen-Wilson & Tyler 1980). This means that an exhaustive analysis is carried out for each new input word, unlike a strictly head-driven parser such as Pritchett’s model (Pritchett 1991) or constraint-satisfaction models (MacDonald, Pearlmutter...
& Seidenberg 1994; Trueswell, Tanenhaus & Garnsey 1994) whose syntactic analyses are contingent upon the heavy use of head information. An incremental parse is comprised of four operations: Match, Inherit, Capture and Submit.

Match is fundamentally a low-level matching operation of input and lexeme. This unification is operated by the best fit principle. The parser works because once an input is matched with the most suitable lexeme, the valency information of the input word, which is acquired by inheritance, is clearly specified. As for words with multiple valencies, the resolution in the parse depends on the architecture of the isa hierarchy. For example, consider matching a transitive verb, eat, as an input with the lexeme EAT:tran, which can be organised in three different ways in the geometry of the isa hierarchy, as shown below. Since valency information is based on the speaker's experience, any of these three configurations in (3.23) could be correct, so it is important that the parser should be compatible with all of these.

In the isa hierarchy, the lower a node is, the more specific information it contains. Since the parser attempts to obtain as much information about a parsed word as possible in order to optimise the parsing, the process of match starts from the top node with the most general information - word in our example (3.23) - on the basis of the Best Fit Principle, and proceeds to as low a node as it can.

(3.23a)

NB: wi = input word; • = instance of word; tran = transitive; intr = intransitive; s = subject; o = object
In the first configuration of (3.23a), matching the input Wi involves two steps. First, the parser matches Wi with EAT:intr at the point Wi is being parsed, because all information about Wi available for the parser is that it is an instance of the lexeme EAT and EAT:intr is the fittest candidate in the isa hierarchy. So, the input Wi inherits the subject valency. Second, at a later stage when the object of Wi is identified in the parse, Wi is rematched with the more specific lexeme EAT:tran and inherits the subject and object valencies so as to accommodate the object.

(3.23b)

If EAT:intr is organised as the more specific case of EAT:tran, as in the second configuration, (3.23b), the input is attached to the lexeme EAT:tran and inherits the subject and object valencies at the point at which Wi is parsed. This is because this time EAT:tran is the fittest lexeme with the input in the isa hierarchy. There will be no change to the match when the parse of the object takes place, since Wi has already had the object valency information and has been anticipating the object. But if Wi had had no object, it would eventually have been matched with the lower node, EAT:intr, as well as with EAT:tran.
As to the last configuration, (3.23c), where EAT:intr and EAT:tran equally constitute the sublexemes of EAT, the matching process requires two stages. First, Wi is linked with the lexeme EAT and inherits the subject valency from the verb node when the input Wi is parsed, because the parser cannot choose the correct sublexeme between the two with the available information about Wi. Second, when the object of Wi is in the subsequent parse, the first match of Wi is cancelled and the rematch of Wi and EAT:tran, the fittest lexeme with the object valency, results so as to parse the object successfully. In consequence, Wi inherits the correct valency information about subject and object.

As shown above, it is possible for the grammar to provide for a given input item for choosing correctly alternative lexical entries with distinct word classes and valent patterns, so a system will be of importance for designing an actual working model. Nevertheless, the choice of this system, e.g. between parallel, serial and delayed processing, is independent of the effects of memory load on the parsing, so I shall leave the issue concerning the treatment of ambiguity open.

Once an input is identified as an example of some lexeme, it receives the information about its lexical entry from the grammar network. Inheritance of appropriate valency information is followed by linking incomplete dependencies in accordance with this information. The best fit principle again performs the linking along with the interaction of the universal principles. It would therefore be true to say that the best fit principle
underlies the whole structure-building operation of parsing. The linking operation is divided into two stages.

The first stage is Capture. This is a search operation for the dependents of the input word. The parser considers every available word that appeared previously to the input to see if it can be its dependent. By available word, I mean any dangling word which has not yet been given a parent. Consider the following situation where \( w_i \) is the current input, \( w_x, w_y \) and \( w_z \) are forming a dependency chain, and \( w_x \) and \( w_p \) are the available words. It is diagrammatically shown below:

![Diagram](image)

The Capture operation on \( w_i \) starts with the most recent dangling word from the input \( w_i \), namely \( w_p \) in our example. If Capture succeeds, the process will be repeated until potential dependents of \( w_i \) are exhausted. Whenever Capture fails, then any further operations should be aborted. Suppose that the first Capture of \( w_p \) failed and the next one of \( w_x \) succeeded in the example. These faulty operations would provoke the illicit dangling of \( w_p \). Given the successful Capture of \( w_p \), the next Capture, in our example \( w_x \), will be consulted.

The second process Submit follows Capture. This time the parser tries to find the parent of the input. As opposed to Capture which can take place recursively, the process of Submit is, as it were, a one-off operation because a word can only have one parent owing to the no-dangling and competition principles. The parser probes into potential parent words preceding the input \( w_i \) in order to establish a relevant dependency. (Obviously there is no need for Submit for a sentence-root.)

Consider a similar example to the previous one with the available word \( w_p \).
appearing before \(w_x\). Under the circumstances, which are illustrated by the figure below, the operation sets off with the most adjacent potential parent to \(w_i\), viz. \(w_z\), i.e. the most recent word which is not subordinated to \(w_i\). The successful Submit terminates the operation. If Submit to \(w_z\) fails, then the process will continue with the next potential parent \(w_y\), which is \(w_z\)'s parent, and so on recursively. All other words (the dots between \(w_x\) and \(w_y\), and between \(w_y\) and \(w_z\)) are inaccessible to the parser for the same reason stated above. Submit terminates after considering the nearest dangling word to \(w_i\), in this example specifically \(w_x\) given that Submit to \(w_y\) has failed. Even though Submit to \(w_x\) failed, no further Submit takes place; there is no need to consider \(w_p\). In fact, whenever the parser encounters a dangling word and either Capture or Submit fails, no more search will be executed. This is because an unfinished vertical arrow, which is assumed to be infinitely long, serves as a barrier, figuratively speaking, that makes its left-hand side area invisible to the parser. By repeating these two processes of Capture and Submit for every input, the parser completes the linking operation without tangling dependencies nor dangling words provided an input string is syntactically well-formed.

\[(3.25)\]

\[
\begin{array}{c}
W_p \quad \downarrow \\
\cdot \quad \cdot \\
W_x \quad \cdot \quad \cdot \\
\downarrow \quad \downarrow \\
W_y \quad \cdot \quad \cdot \\
\downarrow \quad \downarrow \\
W_z \quad \cdot \quad \cdot \\
\downarrow \quad \downarrow \\
W_i
\end{array}
\]

Submit on \(w_i\)

The parse for the input \(w_i\) operates on two data storage spaces, i.e. a stack and a current buffer. The former is a list of data which is structured in such a way that the most recent item before \(w_i\) can be accessed first. This means that the search on Capture and Submit is executed in the "last in-first out" manner: the checking takes place from the tail end of the list. On the other hand, the latter functions like a workbench, so to speak. A word recovered from the stack is transferred onto the current buffer and may be joined with \(w_i\). It is often the case that a stack is depicted as a tray on which items of data are layered vertically. Data retrieval is made at the top of the stack by popping an item off, and processed words are stored by pushing
this onto the stack. The earliest item of data is therefore at the bottom of the stack. A successful parse results in a single dependency chain found on the stack: a single sentence structure consisting of pair-wise dependencies interlinking words for an input string. Thus this model of parsing works just the same as a shift-reduce processor. I assume a single parse stack with unconstrained retrieval of data for our parsing algorithm.

In reality, the parser is unable to have free access to the data in a stack since our working memory degrades over time. One of the classic approaches to this issue is to assume some kind of short-term memory with limited capacity, e.g. $7 \pm 2$ stack cells for Push-down automaton (Yngve, 1960). The problem however, is more serious than simply stipulating constants for limited resource, because the postulated structures of short-term memory such as stacks often lack psychological reality. I shall come back to this matter later, when I treat processing difficulty on multiple centre embeddings in which short-term memory gives rise to severe effects.

Let us observe the incremental steps of the parse for the following sentence: The warm air seems to blow upon her white soft bed. Its full dependency analysis is given below.

(3.26)

The parser takes the following steps to build the dependency structure for the sentence above. The bottom of the stack is indicated by the L-shaped figure, and saved words are put horizontally from left to right onto the stack in the diagrams below. Pop and Push apply not only to an individual word but also to a chunk of words linked by dependencies. For clarity, temporary names such as w1 and w2 are replaced by plain words at the Inheritance stage.
[1-1] READ w1

[1-2] MATCH w1

[1-3] INHERIT w1

[1-4] PUSH w1

[2-1] READ w2

[2-2] MATCH w2

[2-3] INHERIT w2


[3-1] READ w3

[3-2] MATCH w3

[3-3] INHERIT w3
[3-4] POP w2; [3-5] CAPTURE w3

[3-6] POP w1; [3-7] CAPTURE w3: fail

[3-8] SUBMIT w3

[3-9] PUSH w1-w3

[4-1] READ w4; [4-2] MATCH w4

[4-3] INHERIT w4

[4-4] POP w1-w3; [4-5] CAPTURE w4

[4-6] SUBMIT w4: fail; [4-7] PUSH w1-w4

95
The warm air seems to stack buffer.

[5-1] READ w5; [5-2] MATCH w5

[5-3] INHERIT w5

[5-4] POP w1-w4; [5-5] CAPTURE w5: fail

[5-6] SUBMIT w5
The warm air seems to blow.

[5-7] PUSH w1-w5; [6-1] READ w6; [6-2] MATCH w6

[6-3] INHERIT w6

[6-4] POP w1-w5; [6-5] CAPTURE w6: fail

[6-6] SUBMIT w6
The warm air seems to blow upon
The warm air seems to blow upon her.
The warm air seems to blow upon her white.


[9-7] PUSH w1-w8, w9; [10-1] READ w10; [10-2] MATCH w10

[10-3] INHERIT w10


100
The warm air seems to blow upon her


[11-3] INHERIT w11


101
3.3 Dependency Distance and Density

A theory of processing based on Word Grammar offers two methods of measuring syntactic complexity: dependency distance and dependency density (Hudson, 1995b, 1997a). How these two complementary measures work is explained in this section. We shall also see if they correctly predict relative difficulties due to different syntax by applying them to some sentences whose syntactic constructions are well-known for causing processing difficulties.

Processing difficulty can result exclusively from syntactic structure. To illustrate this point, the following pairs of sentences are given below; each pair conveys the same meaning with nearly the same number of words, yet has different syntactic structures. The first member of each pair is harder to process than the second.
(3.27a) That I know about myself as much as when I was born is frightening.

(3.27b) It is frightening that I know about myself as much as when I was born.

(3.28a) I had someone that would follow me to the ends of the earth once.

(3.28b) I had once someone that would follow me to the ends of the earth.

(3.29a) #Castles that mermaids who narcissi surrounded built melted into the sea.

(3.29b) Castles melted into the sea that mermaids built who narcissi surrounded.

N.B. # indicates an unprocessable sentence.

The three pairs contain typical syntactic constructions that can induce processing predicaments. The long subject phrase (3.27a) causes a severe processing load which can be reduced by employing extraposition as in (3.27b). The relative distance between a time adjunct and its parent may be the source for processing difficulty; the longer the distance is as in (3.28), the heavier the processing load can be. The last pair is an instance of centre-embedding (3.29a) and its extraposed counterpart (3.29b). Evidently the relative processing difficulty for these sentences stems purely from their syntactic structures.

What seems to contribute to processing difficulty at the point of processing the input $w_i$ is the number of words that separate $w_i$ from its parent or dependents and the number of words which still have not yet been attached to their parents, bearing in mind that the parser's task for $w_i$ is to find its dependents (Capture) and its parent (Submit). This implies that for assessing syntactic difficulty, there are two parameters, both of which only involve surface structure because extra dependencies virtually follow from surface ones:

(i) dependency distance - the distance between a word and its parent.

(ii) dependency density - the number of incomplete dependencies at any given point of time.

Dependency distance is an indication of the decay of working memory. Because of its limited capacity for inputs, the lifespan of a word in short-term memory is momentary and it gradually fades away from memory. In theory,
the earlier items of data before an input merely become harder and harder to retrieve.\(^1\) It can be stated in dependency terms as follows:

**Principle of Proximity**

The more words are interposed between a word and its parent, the more complex it is to process.

The dependency distance of a given word \(w_x\) is calculated as the number of words between \(w_x\) and its parent. (We shall introduce a more sophisticated measure for Japanese in Chapter 6.) The mean-distance for a text \(T\) is the average value for all words in that text. The following notations are used for these: \(\text{dis}(w_x) = \text{the distance of } w_x; \text{m-dis}(T) = \text{the mean-distance of } T.\) A lower value of \(\text{dis}(T)\) suggests that \(T\) contains simpler and more readable sentences. An illustration (3.30) below is a simple example of dependency analysis with dependency distance, which is shown as a superscript to each word.

\[
\begin{align*}
\text{dis}(w_x) &= \text{the distance of } w_x; \\
\text{m-dis}(T) &= \text{the mean-distance of } T.
\end{align*}
\]

\[
\text{n:s} ^0 \quad \text{v:s} ^0 \quad \text{a:s} ^0 \quad \text{s:a} ^1 \quad \text{v:s} ^0 \quad \text{n:s} ^0 \quad \text{n:s} ^0 \quad \text{v:s} ^0 \quad \text{n:s} ^0 \quad \text{n:s} ^0 \quad \text{n:s} ^0 \\
\text{m-dis}(S_1) &= 0 + 0 + 0 + 1 + 2 + 0 + 0 + 1 / 8 = 0.5 \\
\text{N.B. the distance of the sentence-root is counted as zero.}
\]

The second possible scale for measuring difficulty is dependency density. This displays the number of unattached dependencies at a given point of time. (Accordingly, dependency density is equivalent in spirit to the node-counting measures in phrase structure.) Any incomplete dependencies

\(^1\)It is interesting to note that the immediate recall test of 40 sentences each of which contains 15-21 words, which was conducted by Butterworth, Campbell & Howard (1986), reports significant serial position effects, with the first two words and the last word in a sentence recalled better than the others.
before the input must be accessible for the parser to attempt to Capture or Submit them, which means they are taking up the space of working memory for keeping them active. Once they are linked, the memory load they are placing is assumed to be discharged, because complete dependencies are of no value for the parser. How difficult it is to process a sentence owing to outstanding incomplete dependencies is formalised as below:

**PRINCIPLE OF SPARSITY**

The more dependencies remain incomplete, the more complex it is to process.

The simplest way to measure density would be to treat each open dependency equally and to count the number of incomplete dependencies after each word. It should however, be noticed that a post-adjunct dependency adds nothing to density. To illustrate this matter, let us refer to the preceding example. When the fourth word sad has been parsed, the parse up to this point is, as it were, complete in the sense that no word is specially expected. (I am only sad is a well-formed sentence by itself.) Hence the density after sad has been processed is zero. On the other hand after processing the subsequent word in, the density increases to one due to the anticipation of its complement. Dependency density is exhibited by the following notation:

dens(wx) = the density immediately after wx; m-dens(T) = the mean-density of T. A higher dens(wx) shows that the parser is under pressure due to heavy memory load when wx is being parsed. The figure (3.31) below demonstrates the dependency density for the previous example, which is displayed as a subscript.

![Diagram](image)

\[ m\text{-dens}(S_1) = \frac{1+1+2+0+1+1+2+0}{8} = 1.0 \]
We now turn to the preceding three pairs of sentences (3.27), (3.28) and (3.29), and actually measure their dependency distance and density to examine whether the relative processing difficulty of the pairs is manifested in the evaluation. First the dependency analysis with the distance and density for the three pairs are given, then it is followed by Table 3.1 of their mean-distances and densities.

(3.27a)

(3.27b)

(3.28a)
had once someone that would follow me to the ends of the earth

CASTLES that mermaids who narcissi surrounded built melted into the sea

had, once, someone that would follow me to the ends of the earth
Table 3.1 Mean-distances and densities for (3.27), (3.28) and (3.29)

<table>
<thead>
<tr>
<th>sentence</th>
<th>number of words</th>
<th>gross distance</th>
<th>mean distance</th>
<th>longest distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.27a</td>
<td>14</td>
<td>16</td>
<td>1.14</td>
<td>11</td>
</tr>
<tr>
<td>3.27b</td>
<td>15</td>
<td>6</td>
<td>0.40</td>
<td>3</td>
</tr>
<tr>
<td>3.28a</td>
<td>14</td>
<td>12</td>
<td>0.86</td>
<td>11</td>
</tr>
<tr>
<td>3.28b</td>
<td>14</td>
<td>2</td>
<td>0.14</td>
<td>1</td>
</tr>
<tr>
<td>3.29a</td>
<td>11</td>
<td>14</td>
<td>1.27</td>
<td>6</td>
</tr>
<tr>
<td>3.29b</td>
<td>11</td>
<td>5</td>
<td>0.45</td>
<td>3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>sentence</th>
<th>number of words</th>
<th>gross density</th>
<th>mean density</th>
<th>highest density</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.27a</td>
<td>14</td>
<td>24</td>
<td>1.71</td>
<td>3</td>
</tr>
<tr>
<td>3.27b</td>
<td>15</td>
<td>13</td>
<td>0.87</td>
<td>2</td>
</tr>
<tr>
<td>3.28a</td>
<td>14</td>
<td>9</td>
<td>0.64</td>
<td>1</td>
</tr>
<tr>
<td>3.28b</td>
<td>14</td>
<td>10</td>
<td>0.71</td>
<td>1</td>
</tr>
<tr>
<td>3.29a</td>
<td>11</td>
<td>21</td>
<td>1.91</td>
<td>5</td>
</tr>
<tr>
<td>3.29b</td>
<td>11</td>
<td>9</td>
<td>0.82</td>
<td>2</td>
</tr>
</tbody>
</table>

Comparing the set of the easier sentences to process, namely (3.27b), (3.28b) and (3.29b) with the other set consisting of the harder sentences, viz. (3.27a), (3.28a), and (3.29a), we can readily see the striking difference in their mean distances. The latter set is more than twice as great as the former set. Dependency distance therefore, makes the correct prediction for relative processing difficulty. Interestingly the fact that (3.27b) and (3.29b) are the extraposed versions of their counterparts reinforces the view that extra dependency has virtually no effect on difficulty; besides extraposition appears to reduce difficulty in spite of increasing complexity. As for the longest distance, it appears to have nothing to tell us about processability.

Nonetheless, dependency density apparently shows a different aspect of processing difficulty from the one that dependency distance presents. Mean density as opposed to mean distance fails to predict correctly the difference of complexity between (3.28a) and (3.28b); the latter which is supposed to be easier to process bears a higher mean density than the former. On the other hand it does predict the relative difficulty of the other pairs. If density monitors the size of temporal memory load at any given time, there is a strong connection between the uprocessability of (3.29a) and its maximum density with the figure of 5, which is the most extreme in comparison to
the rest. Density thus has the advantage over distance of showing where the memory demands are highest within a sentence.

In conclusion, dependency distance and density display separate aspects of processing difficulty. Measuring the mean distance of a given text would be the best approximation to its overall syntactic difficulty or readability. On the other hand, to judge the processability of sentences, the highest density might have to be referred to. I shall come back to this subject later in connection with processing breakdown and memory limitations.

Finally, we will consider the two questions raised at the beginning of the chapter: (i) What, if any, are the fundamental differences between a parser assuming phrase-structure grammar and a Word Grammar-based parser? (ii) Which model of sentence-structure should be preferred from the parsing viewpoint?

There is at least one main difference between the phrase structure-based and the Word Grammar-based parsers. The relationships between words are expressed differently; while the relations are shown by higher non-terminal nodes in phrase structure, a classified dependency between the words shows them directly in Word Grammar. This immediately indicates two advantages of a Word Grammar-based parser for theories of parsing.

Firstly since most non-terminal nodes in a phrase-structure analysis are basically copies of the features associated with a head, the parser has to maintain large amounts of redundant information. In contrast, each word in the Word Grammar model is linked with only one parent, so the syntactic information on each word is maintained more economically by the parser.

Secondly the phrase-structure model offers an extensive range of alternative analyses of a given sentence varying from grammar to grammar, each of which will give different measurements of syntactic complexity. In contrast, dependency analysis allows far fewer options for alternative analyses.

As an extra consequence of the difference in the parser's task, there arises a technical difference as to the parse of post dependents, where a phrase structure-based parser may have to take more complicated steps. Parsing a post dependent involves grafting a new node into the middle of the
constructed tree, which necessitates back-tracking and undoing parts of the tree. This is because the parser is always forced to choose a certain phrase structure at the point any given word is parsed. Take the following sentence (3.30) for example.

(3.30) I eat slowly.

For the phrase structure model to parse the post adjunct word, slowly, in (3.30), the parser has to undo the VP node governing the verb, eat, and build the higher VP node in order to accommodate the AP that governs the adverb, slowly. Consequently the newly constructed VP has to be implanted in the existing tree, so (3.31) below demonstrates the parsing steps for the input slowly.

(3.31)

In contrast, the parse of the post adjunct, slowly, in the Word Grammar model is less complicated and completed by merely linking it with its parent word, eat, without changing the existing dependency structure, as shown in (3.32) below. This is due to the fact that a Word Grammar parser aims at finding only one head for each word in the parse (except for the sentence-root verb).
If the economy of syntactic parsing is of vital importance to a parser, the Word Grammar model can offer the more parsimonious course of parsing, therefore it should be preferred to the phrase structure model. This is not only because of the handling of post dependent mentioned above, but is related to the simplicity of the nature of parsing in the Word Grammar model, which is ultimately satisfaction of valency requirements specified in grammar. By contrast, because phrase structure assumes a division between lexicon and phrase-structure rules or more general principles, the parser has to take two major steps to construct a phrase structure for words in the parse: generating structures on the basis of the rules or principles and projection from the lexicon.
4 Words in Word Grammar

In this chapter, firstly we recapitulate what has been established about words in Word Grammar. Secondly, we attempt to verify the status of the word for some problematic cases such as clitics and compounds in English and Japanese. The discussion will also consider how much their wordhood matters in respect of parsing. Lastly I shall offer a tentative list of word-classes for Japanese.

4.1 The word in Word Grammar syntax

As explained previously, Word Grammar maintains that words are the maximal units of syntax; any larger units such as phrases are redundant. The grammar treats the structure of an entire sentence as a set of pairwise relations amongst its words. There are, however some exceptions to this. In handling coordination and direct speech as in the examples below, the grammar refers to a word-string.

(4.1) They are a broken skull and a leather cosh.
(4.2) He said, "For liberty there is a cost."

To complete a syntactic analysis for these examples, the grammar ought to recognise word-strings: a broken skull and a leather cosh and For liberty there is a cost in (4.1) and (4.2) respectively. It is worth emphasising that the notion ‘phrase’ is still not needed in Word Grammar because the grammar can be defined without making any reference to phrases.

Words, on the other hand, are also the minimal elements of syntactic structure in Word Grammar syntax, which follows the tradition of Word-and-Paradigm morphology (Robins 1959, Matthews 1972). This approach takes the view that morphological structure is invisible to syntax. Morphology therefore has charge of the internal structures of words; syntax, in contrast to it, governs external relations among words. Word Grammar again allows exceptions where one word is made up of other words, as in the case of cliticisation and compounds. We shall consider this point in the next subsection.
4.2 Syntactic words versus phonological words

So far it has been understood that words play a central role in Word Grammar syntax. Yet, one might query how a word can be recognised without a clear definition of it. As we have observed, Word Grammar is a network theory in which concepts are interconnected with one another, with an ISA hierarchy being one major sub-network. Within this organisation, any concepts including the concept 'word' are characterised solely by their typical properties, any of which (thanks to default inheritance) can be overridden for handling exceptions. For instance, properties of the typical word would readily be listed: it has a sound pattern defined by normal phonology, it has the syntactic function of a dependent, and it has a meaning; but all of these properties allow exceptions - untypical phonological patterns in loan words, words which require no parent, and words without ordinary meaning. Hence the concept 'word' could never be defined by articulating its necessary and sufficient conditions, nor could any other concepts.

So far as syntax is concerned, then, words are the essential units of syntactic structure. Furthermore these same words should also be the essential units of morphology, bearing in mind that they are the link between syntax and morphology. The largest unit in morphology is a word which in turn is the smallest unit in syntax. In addition to the word in syntax and morphology, it is however necessary to recognise at least another abstract concept of 'word' in the phonological sense. (I shall ignore the word in semantics since it is not directly related to the research topic.) We shall call the former '(morpho-)syntactic word' and the latter 'phonological word'. The necessity for recognising two word-types is due to mismatches such as those reviewed below, where syntactic word boundaries do not coincide with phonological ones.

4.2.1 The contracted forms of verbs

In the subsequent sections, we shall examine so-called clitics and compounds in English with regard to their statuses as syntactic words. The first case is the contracted forms of verbs, which are clitics. Take the following sentence containing the contracted form of 'are' as an example.

(4.3) You're sinking.
In phonology, (4.3) consists of two phonological words: /jo:/ and /siqkɪ/, since the first cluster /jo:/ would represent no more than two phonemes if it were divided into the consonant and the vowel. In contrast, there must be recognised three syntactic words. In other words, the clitic, viz. the contracted form of the BE verb should be analysed as an individual word; 're in this case. If the cluster 'you're' is taken as one syntactic word, it will have to be analysed as though it were both sentence-root and subject. Hence, the structure in the right in (4.3) will be syntactically illegal owing to an infringement of The Competition Principle: in surface structure, each word depends on no more than one word. The diagram below will illustrate this point.

![Diagram showing the structure of (4.3)]

4.2.2 The possessive "s"

Along the same line as reduced forms of English verbs, there is another putative clitic: the possessive marker "s". There are two positive motives for treating it as a separate syntactic word. Consider (4.4), (4.5) and (4.6) below.

(4.4) his girlfriend
(4.5) the neighbour's girlfriend
(4.6) the neighbour living on the ground floor's girlfriend

Primarily, comparing the first two examples, we can observe that 'neighbour's' in (4.5) has the function of determiner. It is therefore possible to regard the cluster in question as a compound of a common noun and a
pronoun. Nevertheless, this analysis is implausible since it fails to show a dependency relationship between the preceding noun and 's' in such group genitive cases as (4.6). Notice that the possessor is not 'floor' but 'the neighbour' in this example. Thus the possessive marker 's' has the status of a syntactic word. The following diagram is the dependency analysis for (4.6), with two syntactic words corresponding to the single phonological word /flɔːz/.

(4.6)

Thus far, the evidence has shown clearly that such clitics as the contracted forms of verbs and the possessive 's' in English are independent syntactic words. In the following subsection, we shall investigate the rather ambiguous case of a noun-noun pair that could be seen as either a compound constituting one syntactic word or a noun plus noun sequence comprised of two syntactic words.

4.2.3 Noun plus noun pairs

The noun plus noun pairs include compound nouns such as 'black sheep' and 'blackmail'. Since some such pairs are written with or without a word space, they sanction two syntactic analyses: one syntactic word of a compound or two syntactic words composed of a noun-noun sequence. The pair 'ground floor' in (4.6) may well be a noun plus noun compound especially when it depends on another noun; its spelling with a hyphen reinforces this assumption, e.g. a ground-floor flat. Moreover, occasionally there is an unclear case in which applying dependencies to the nouns within a pair is extremely awkward. For example, the dependency relation between 'Apple Macintosh' is ambiguous since both are the registered trademarks and can be used independently. In the absence of sound motivation for recognising two syntactic entities in this type of pair, we are left with two acceptable
syntactic analyses for it. Below are given two alternative dependency analyses for 'ground floor'.

\[(4.7a)\] \[(4.7b)\]

In terms of syntactic difficulty, what counts is not how noun plus noun pairs are syntactically processed but how much processing load they carry. Let us turn to the above examples, and assume that the processing loads for the words 'ground', 'floor', and 'ground-floor' are \(x_1\), \(x_2\), and \(x_3\), respectively. The sum of \(x_1\) and \(x_2\) in (4.7a) should equal \(x_3\) in (4.7b), because both are the identical string of words. Thus, provided that the same amount of processing load is assigned to both (4.7a) and (4.7b), the exact syntactic structure for 'ground floor' would not be of significance in processing difficulty.

Ultimately, the way compounds are processed would be determined by how they are related to their permanent words in the isa hierarchy. If one has a node of sub-lexeme GROUND-FLOOR, then (4.7b) will be the more probable manner for processing it, or else 'ground' and 'floor' will be processed as two separate words. How the noun-noun pairs are preserved in our brains may be different from one person to another, and from one pair to another. It is therefore, worth allowing alternative syntactic analyses for these pairs, while finding some other basis, independent of the syntactic analysis, for predicting processing difficulty.
Two ways of accessing the stored words for the utterance ground floor. (4.7c) and (4.7d) correspond to the resulting syntactic structures of (4.7a) and (4.7b) respectively.

4.3 Syntactic wordhood in Japanese

In terms of identifying words, Japanese is a particularly problematic language. This is mainly due to two reasons. Firstly, as pointed by Tsujimura (1996: 125), word spaces are not institutionalised in the writing system. Secondly, the language displays a high degree of agglutination, and thus is classified as an agglutinative language in the domain of language typology. The linear sequence (4.8) demonstrates this phenomenon. In spite of its length and internal complexity, this sequence would count as a single syntactic word if our only criterion was freedom of order because the order of morphemes is rigidly fixed.

(4.8) ikasaseraretakuwaarimasendeshitadeshou.
I bet you did not want to be made to go.

The string consists of twenty graphemes in Japanese orthography and can be decomposed into twelve morphs. Below, (4.8) is rewritten with spaces for
grapheme boundaries and larger spaces indicating morpheme boundaries.

\[(\text{4.8'}) \text{i ka sa ra re ta ku wa a ri ma se n de shi ta de shou.}\]

to go causative passive to want topic to be polite negative copula past copula future
I bet you did not want to be made to go.

The following discussion considers the syntactic wordhood of such unclear cases as particles, auxiliary verbs and counters in order to establish a basis for a consistent dependency analysis of Japanese sentences and also for measuring their syntactic difficulty, which will be important in the succeeding chapters. Although there is previous work in the framework of Tesnière's structural syntax such as Lepage, Ando, Akamine & Iida (1998) that crudely classifies Japanese words into two classes, i.e. content words and function words (i.e. particles), the word-status of auxiliary verbs and counters is not discussed. Further, at the end of this chapter I shall raise the question of the smallest processing unit in connection with this discussion of syntactic words.

4.3.1 Particles

Numerous attempts have been made in the linguistics literature to resolve the inconclusive status of Japanese particles (Hattori 1960; Suzuki 1972; Martin 1975; Sakakura 1980; Oyano 1982; Vance nd). The succinct account by Zwicky (1977: 1) addresses the problem apropos of the particles as follows:

Most languages - very possibly all, except for the most rigidly isolating type - have morphemes that present analytic difficulties because they are neither clearly independent words nor clearly affixes.

In the phrase-structure syntax framework (Harada 1977; Saito 1985), the particles are regularly treated as affixes on the grounds that some particles allegedly behave like grammatical case markers. I shall begin the discussion with this type of particles.

The particles 'ga', 'o', 'no' and 'ni' are claimed to be the typical grammatical cases, i.e. the nominative case, the accusative case, the genitive case, and the dative case, respectively. ('ni' will be dealt with later in this section.)
Thanks to the indication of a noun’s function by a following particle, word-order in a sentence or phrase is relatively free, subject only to the general restriction that a dependent precedes its parent. Compare the following examples of (4.9) through (4.14). Notice that an adjective sharing the same parent as an alleged genitive noun can be interposed between the two nouns as shown in (4.14). However, as expected, tangled dependencies (which produce discontinuous phrases) are not allowed; cf. (4.12).

(4.9) boku ga kimi no koohii o nomu. (I drink your coffee.)
     I your coffee to drink

(4.10) kimi no koohii o boku ga nomu.
       your coffee I to drink

(4.11) *boku ga koohii o kimi no nomu.
       I coffee your to drink

(4.12) *kimi no boku ga koohii o nomu.
       your I coffee to drink.

(4.13) oishii kimi no koohii o nomu. (I drink your delicious coffee.)
       delicious your coffee to drink

(4.14) kimi no oishii koohii o nomu.
       your delicious coffee to drink

So far these particles seem to act as though they were case marking the preceding nouns. Irrespective of the resemblances, there are some striking differences between the particles and typical case inflections. Firstly, some are optional. Accordingly, the omission of such particles as ‘ga’ and ‘o’ has no effect on grammaticality, although occasionally omitting particles can cause syntactic ambiguity. Compare the above example (4.9) and its version without ‘ga’ and ‘o’ (4.15), which is grammatical.

(4.9) boku ga kimi no koohii o nomu. (I drink your coffee.)
     I your coffee drink
Secondly, the alleged case particles do not mark their functions uniquely. Take the particle ‘ga’ for instance. Despite the fact that this particle should generally regularly serve as the subject marker, it generally specifies the object of a stative verb (Tsujimura 1996: 211), as seen in (4.16). (Notice that the alleged dative case ‘ni’ can also mark the subject in this construction.) Furthermore, (4.17) shows that within a relative clause ‘ga’ can be converted to ‘no’ when it marks the direct object of a stative verb (Harada 1971, 1976b; Miyagawa: 1993).

(4.16) Jako ni/ga Huramango ga wakaru. (Jaco understands Flemish.)

Jaco Flemish understand

(4.17) Huramango no/ga wakaru Jako ga ryouri ga jouzu da.

Flemish understand Jaco cooking good copula

(Jaco who understands Flemish is good at cooking.)

Furthermore, the complement of this putative nominative case is not restricted to a noun but can be a verb as in (4.18).

(4.18) iku ga yoi. (You had better go.)

go good

Lastly, nouns cannot be marked by the alleged case particles in coordination as opposed to those in a case language where case is always used to manifest the function of each coordinated noun. Let us examine the Japanese coordinated sentences (4.19), (4.20) and (4.21) and the Russian equivalence (4.22) below.

(4.19) Boku wa mizu to uokka o nomu. (I drink water and vodka.)

I water and vodka to drink

(4.20) Boku wa mizu to uokka to o nomu.

I water and vodka and to drink

120
Comparing (4.19) and (4.22), we can clearly see that a Japanese string of a noun plus the particle 'o' is quite dissimilar to a Russian case-inflected noun; in Japanese a noun and its following particle are not a single word composed of a stem and a suffix, but rather separate units. The Russian equivalent sentence (4.22), in contrast, reveals that two accusative nouns 'vodu' and 'vodku' are coordinated; cf. 'voda' and 'vodka', which are the corresponding nouns bearing nominative case, respectively. Besides, as can be observed from (4.19) and (4.20), the particle 'o' is a syntactically independent word, since there must be a dependency relationship between 'o' and its preceding coordinating particle 'to'.

The following conclusions can therefore, be drawn: Japanese particles are clitics in the sense that although they are appended to free-standing words, they are separate syntactic words. In addition, although some particles are utilised to mark the grammatical function of the preceding word, they are not conventional case markers. Figure (4.23) is the dependency analysis for (4.19) illustrating the treatment of particles in the subsequent chapters.

A lexicalist approach to Japanese morpho-syntax such as Sells (1995; 1996) is analogous to that of Word Grammar syntax in that internal morphological structure is opaque to syntax, but case-marking particles are treated as suffixes. For instance, the string of a noun and a particle 'boku wa' is
analysed as a lexical morpheme 'boku' plus a nominal suffix 'wa'. The following diagram for 'boku wa' is presented according to the lexical analysis by Sells (1996).

\[
\text{(4.24)} \quad \begin{array}{c}
\text{NP} \\
\text{N} \\
\text{N}^0 \\
boku \\
\text{wa}
\end{array}
\]

The convention used in (4.24) is that 'boku' as lexicon-internal category is represented by N while N^0, in this case 'boku-wa', is a single syntactic category as a word, which is to be inserted in the syntax. Consequently the syntax has no access to the internal structure below N^0. Although the Lexicalist handling of syntactic words in Japanese differs from that of Word Grammar illustrated above, the question of which syntactic structure is more plausible will be unimportant in terms of parsing so long as the same amount of processing load is assigned to the whole structure.

Other kinds of particles will be briefly examined here. One is a group of particles that identify the semantic functions of their dependents. They have been dealt with as either oblique cases or postpositional suffixes in the mainstream of linguistics (Tsujimura 1996). In contrast I shall again treat this group of particles as syntactically independent words, for they bear the same characteristics as the alleged grammatical case particles. The particle 'ni' is a typical example; it has several semantic functions in spite of its inherent implication of direction and the word-class of its dependents is not restricted to noun.

\[
\text{(4.25)} \quad \text{Oberisuku wa konkorudo hiroba ni aru.}
\]

\[
\begin{align*}
\text{Obélisque} & \quad \text{Concorde} \\
\text{place} & \quad \text{in} \\
\text{to be}
\end{align*}
\]

(The Obélisque is in the Place de la Concorde.)

122
(4.26) Pari ni iku. (I am going to Paris.)  
Paris to to go

(4.27) Ane ni ryouri o oshieru. (I teach cooking to my sister.)  
sister to cooking to teach

(4.28) Ane ni ryouri o osowaru. (I learn cooking from my sister.)  
sister from cooking to learn

(4.29) wain ni chiizu. (wine and cheese)  
wine and cheese

(4.30) Naku ni nake nai. (I cannot possibly cry.)  
to cry to cry not

(4.31) Machi ni mata. (I waited and waited.)  
to wait and to wait past

The first five examples present various semantic functions of the particle 'ni'. On the other hand, (4.30) and (4.31) show that 'ni' can have not only a noun as its dependent but also an infinitive or a particular inflected form of verb; cf. the infinitive forms of the verbs in (4.30) and (4.31) are 'naku' and 'matsu', respectively. Hence, 'ni' is not a case marker.

Furthermore, this group of particles, like the putative case particles, cannot be repeatedly attached to linked items in coordination. This is illustrated in the following (4.32) and (4.33).

(4.32) Densha to basu de Pari ni iku.  
train and bus by Paris to to go  
(I am going to Paris by train and bus.)

(4.33) *Densha de to basu de Pari ni iku.  
train by and bus by Paris to to go

---

1 'ni' in the sense of direction like the one in this example can be omitted in casual speech, but this is atypical of postpositional particles in general.
It should be remembered that the particles 'wa', 'ga' and 'o' and the group of particles functioning as postpositions that have been described here form different sub-divisions within particles. For one thing, the former has no particular semantic content but its role is to specify the grammatical function of its dependent, whereas the latter bears an inherent meaning, e.g. 'kara' (from) indicates a starting point and 'made' (up to, until) shows a limit or a boundary. Besides, the omission of the former is allowed in casual speech, whereas leaving out the particles in the latter group gives rise to syntactic ill-formedness. Compare the sentences in (4.34) below. In addition, the former type of particles can take another particle as a complement as shown in (4.35). A distinction should therefore be drawn between the particles bearing syntactic functions and the postpositional particles.

(4.34) Kanojo wa yubi o niahu de kit ta.

she finger knife with to cut past
(He cut her finger off with the knife.)

Kanojo yubi naihu de kit ta.
she finger knife with to cut past

*Kanojo yubi naihu kit ta.
she finger knife to cut past

(4.35) Jounai de wa tabako wa kinen desu.
inside the hall in cigarette no smoking copula
(Smoking is prohibited in the hall.)

Nevertheless, some particles seem to be associated with both categories. For instance, 'mo' (also), 'no' (the possessive marker) are primarily postpositional particles, thus their presence is obligatory to retain their meanings. It is however possible for 'no' to specify a subject in a relative clause construction, as observed in (4.17). Similarly, (4.36) suggests that 'mo' too can serve as a grammatical function marker. Also these particles are able to be combined with other particles, e.g. (4.37). 'mo' and 'no' therefore belong to both groups of functional particles and postpositional particles.
(4.36) Watashi mo datsuzei ga shi tai.
I also tax evasion to do desiderative
(I too would like to evade tax.)

Watashi wa datsuzei mo shi tai.
I tax evasion also to do desiderative
(I would like to evade tax too.)

(4.37) Jounai de no otabako wa o-hikae kudasai.
inside the hall in of cigarette refraining to give
(Please refrain from smoking in the hall.)

Watashi wa datsuzei o mo kokoromi ta.
I tax evasion also to attempt past
(I attempted tax evasion too.)

Discourse particles form another group. They do not express either syntactic or semantic relationships but convey the speaker's attitude and emotions. It is without exception possible to attach these particles at the end of phonological units where there could be a pause (i.e. phonological words - see earlier discussion). This distinction is easy to explain if they are adjuncts, i.e. syntactic words, but very hard if they are inflections. They must however, be post adjuncts, which is against the general syntactic rule in Japanese that any word is pre-dependent of the parent. (4.38) demonstrates some examples of the discourse particles and their dependency structures, where 'ne', 'sa', and 'yo' mean something extra such as '......, you see'.

(4.38)
The final group of particles link nouns, clauses, and sentences such as 'to' (and), 'node' (because) and 'ga' (but). It would be uncontroversial to assume that they are separate syntactic words holding a pre-dependent relation with the following dependent words.

4.3.2 Auxiliary verbs and copula

In Japanese, verbs, adjectives and auxiliary verbs including the copula share the property that they are inflected. While verbs and adjectives can stand independently (i.e. as free morphemes), auxiliary verbs can typically occur only as bound morphemes appended to hosts. Some linguists (Shibatani 1990) treat them as suffixes of verbs or adjectives owing to the fact that the inflectional elements of the latter subcategorise for particular auxiliary verbs and particles. It seems however, more logical to regard them as syntactically separate words on account of their being inflected; auxiliary verbs have the same inflections as full verbs. Table 4.1 manifests the standard inflections of a full verb 'tobu' (to fly), an auxiliary verb 'masu' (polite) and some of their subcategorised auxiliary verbs and particles.

Table 4.1 Inflections of a full verb 'tobu' (to fly) and an auxiliary verb 'masu' (polite)

<table>
<thead>
<tr>
<th></th>
<th>to fly</th>
<th>polite</th>
<th>subcategorised auxiliaries and particles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Irrealis</td>
<td>tob-a</td>
<td>mas-e</td>
<td>nu, na-i (Negative), se-ru, sase-ru (Causative)</td>
</tr>
<tr>
<td>Adverbial</td>
<td>tob-i</td>
<td>mash-i</td>
<td>ta (Past), ma-su (Polite), ta-i (Desiderative)</td>
</tr>
<tr>
<td>Conclusive</td>
<td>tob-u</td>
<td>mas-u(ru)</td>
<td>rash-i (Hearsay), to, ga, kara (Particle)</td>
</tr>
<tr>
<td>Attributive</td>
<td>tob-u</td>
<td>mas-u(ru)</td>
<td>-</td>
</tr>
<tr>
<td>Hypothetical</td>
<td>tob-e</td>
<td>mas-ure</td>
<td>ba (Particle)</td>
</tr>
<tr>
<td>Imperative</td>
<td>tob-e</td>
<td>mas-e</td>
<td>-</td>
</tr>
<tr>
<td>Cohortative</td>
<td>tob-o</td>
<td>mash-yo</td>
<td>u (Particle)</td>
</tr>
</tbody>
</table>

N.B. - indicates a segmental boundary between a stem and an inflectional suffix.

Then we can explain that auxiliary verbs as well as particles are enclitics, which would also explain their rigid word order that otherwise suggests that they must be affixes. The schematic diagram (4.39) below illustrates the arrangement of a verbal phrase and the compositions of a verb and an auxiliary verb.
Having established the syntactic wordhood of an auxiliary verb, we focus attention on one minor point: the auxiliary status of 'u' (future).

The inflection 'cohortative' of verbs followed by the category 'u' yields the future form (Sakuma 1936) and the presumptive form (Bloch 1946) as shown in (4.40).2

(4.40) Ame wa yamo u. (It will stop raining.)

Sakuma and Bloch's analysis of 'u' is that this is a part of the inflectional suffix, i.e. '-ou' and '-yo' are the inflectional suffix of verbs. On the other hand, 'u' is considered as an auxiliary verb in the grammatical tradition. Conversely, a problem also arises with the traditional account for this element, since 'u' has no other inflected forms as opposed to the other auxiliary verbs. The auxiliary status of 'u' is therefore doubtful. Thus I postulate that 'u' along with the negative tentative 'mai' which also never inflects is a particle. Except for the treatments of 'u' and 'mai', I employ the standard view suggested by Shibatani (1990: 232) in connection with the inflectional paradigms of verbs and auxiliary verbs given in Table 4.1.

2 In Modern Japanese, this form more conventionally serves as cohortative (e.g. tobo u "Let us fly" or intentional (e.g. tobo u to omou "I think I will fly").
4.3.3 Numeral quantifiers


(4.41) ichi-zen one bowl of rice roku-wa six birds
ni-ki two receptacles nana-satsu seven bound objects
san-sou three boats hachi-mai eight thin flat objects
yon-hiki four animals kyuu-hai nine cups of beverage
go-tou five heads of cattle juu-jou ten tatami carpets

Numerals are usually expressed with one of ten digits followed by a mathematical term for the order of magnitude as in (4.42). Since such terms as sen (thousand), hyaku (hundred), and juu (ten) are self-standing words and on their own they mean one thousand, one hundred, and ten respectively, these magnitude-indicators and digits ought to be treated as syntactically separate.

(4.42) kyuu man nana sen go hyaku ni juu ichi
nine ten thousand seven thousand five hundred two ten one
(97521)

With respect to classifiers, they are conventionally analysed as suffixes of the numerals that they are attached to on the grounds that they are bound morphemes. However, bearing in mind that there are ten possible combinations with digits for any classifier, it should probably be handled as a separate syntactic word, i.e. one lexeme. Moreover, although numeral quantifiers show some phonological variations, as illustrated in (4.43) by the variations in the classifier ‘hon’, this is predictable by phonological rules.
The other evidence for the treatment of a classifier as an enclitic to a digit or a magnitude-indicator is seen in such coordination structure as (4.44). The dependency analysis of this example tells us that 'gosoku' (five pairs of footwear) is not allowed to be one syntactic word because 'go' (five) must be coordinated with its preceding numerals and magnitude-indicators.

(4.44) 

In consequence, I suggest that a numeral quantifier should be treated as a combination of two independent words: 'digit' and 'classifier'; in the isa hierarchy therefore, numeral quantifier will not be recognized as one lexeme. The figure (4.45) below shows the links between the temporary words $w_1$ plus $w_2$ in 'is soku' (a pair of footwear) and their corresponding lexemes ICHI and SOKU, as part of the isa hierarchy.
4.4 Outline of Japanese Word Grammar

This section offers a brief description of Japanese Word Grammar. I will present the word-classes, dependency types, and inflections of the language in question. Subsection 4.4.4 will then add a summary for notation.

4.4.1 Word-classes

In the Word Grammar theory, word classes are set up according to the principle that they allow generalisations not only of syntax but of morphology, semantics and others which would not otherwise be possible (Hudson 1997b). I will apply this principle in an attempt at grouping Japanese words into their word-classes.

Japanese seems to require four major classes: adverb, noun, verb, and particle, the last three of which are further subdivided. The subdivisions of particle are, as I have pointed out in the previous section, those signalling grammatical information, those functioning as postpositions, those used as
coordinators, and those conveying discourse information. I shall label these sub-classes as function particle, postposition particle, coordination particle and discourse particle, respectively. Concerning verbs, there are three sub-classes, namely, full verb, auxiliary verb and adjective - the reason for this classification will be explained later. Noun contains seven sub-classes; alongside common noun, proper noun and pronoun, which is traditionally termed demonstratives, I shall distinguish four sub-classes idiosyncratic to the Japanese language, viz. numeral noun, classifier, adjectival noun and verbal noun. The previous section has briefly introduced the former three categories. The following discusses the nature of the verbal noun.

As its name connotes, the verbal noun (Martin 1975, Kageyama 1976-7, Iida 1987, Miyagawa 1987, Shibatani and Kageyama 1988) possesses dual properties of both noun and verb. Most members of this class arise from Sino-Japanese and loan words, but some are native Japanese compounds of noun and deverbal noun. (4.46) provides some examples of verbal nouns.

(4.46) Sino-Japanese loan words native Japanese
shuryou 'hunting' iiimeeru 'email' dokumi 'tasting food before serving'
suiei 'swimming' ranningu 'running' yukikaki 'shovelling snow'

Verbal nouns are essentially nouns in the sense that they can be complement of a function particle, and may take a pronoun or adjective as adjunct, as demonstrated in (4.47). They are however, differentiated from typical nouns in that verbal nouns are able to form compound verbs in combination with the full verb 'suru' (to do) as opposed to typical nouns. This is made clear in (4.48).

(4.47) kare no dokumi ga hitsuyou da.
he of tasting food necessity copula
(We need to let him taste the food before serving.)

kono uttoushii yukikaki o suru.
this dismal shovelling snow to do
(I do this dismal snow-shovelling.)

3 Verbal nouns of Japanese origin are not necessarily compound nouns. As a marginal case, ocha 'tea' can be a verbal noun, cf. ochasuru 'to have tea'.

131
dokumisuru ‘to taste food before serving’ *dokusuru (to poison) < doku (poison)
yukikakisuru ‘to shovel snow’ *yukisuru (to snow) < yuki (snow)

The other property of verbal noun that is attributed to verb is expressed in its marking on object. It is normally the case that an object of a verb is specified by the function particle ‘o’. The verbal aspect of a verbal noun is borne out by the example (4.49) where the object of the verbal noun ‘yukikaki’ is identified by ‘o’. Further evidence for the classification of a verbal noun as a verb is that ‘gatera’ in (4.49) is used in depicting simultaneous actions and its complement can only be either a full verb or a verbal noun. Note however, that the object of ‘yukikaki’ in (4.49) will serve as a prenominal modifier when it is a complement of ‘no’. This phenomenon serves as evidence that the verbal noun enjoys dual status of verb and noun. To sum up, verbal noun and adjectival noun are also sub-classes of noun, where multiple default inheritance must be allowed to apply.

(4.49) yane o/no yukikaki gatera yukigassen o shi ta.
roof shovelling snow while snowball fight to do past
(We had a snowball fight while shovelling the snow on the roof.)

For words to which no kind of generalisations could be made, no word-class would be assigned. Such expletives as ‘un’ (yes), ‘uun’ (no), ‘yaa’ (hello) and ‘waa’ (wow) are left unclassified. Since Word Grammar word-classes are structured in an inheritance hierarchy with the topmost category “word”, then these class-less lexemes are positioned immediately under ‘word’. The overall sketch of the word-classes for Japanese is presented in Figure (4.50).
4.4.2 Dependency structures

Next, we will explore the dependency relationships for Japanese in terms of which syntactic knowledge is expressed. These direct relationships between words ought to be considered from two aspects: word order and grammatical function.

The word order between a parent and a dependent for Japanese entails one default type of dependency: 'dependent', which defines the dependent preceding the parent. This is because the arrangements of Japanese words are generally head-final, as has been pointed out in the introduction chapter. This generalisation in turn means that extraposees such as inverted elements following a sentence-root word and discourse particles displayed in (4.51)* should be handled as exceptions thanks to their being post-dependents ('post-adjuncts' or 'extraposees', labelled >a or >x).

---

* Pronouns such as kono ‘this’ in this example are pre-dependent in Japanese unlike those of English, because they are always optional and syntactic well-formedness is never affected by their omission. Also they can either precede or follow any other dependent of the same noun. In addition, it should be noted that there are no other so-called determiners in Japanese.
The other way to classify dependency types is on the basis of grammatical relations between a dependent and a parent. One criterion for dividing them into the major two categories, valent and adjunct, is whether the parent provides a dependency frame. Dependents filling syntactic slots supplied by their parents are defined as valents, on the other hand those not filling such slots are classified as adjuncts (which may simply be identified with 'dependent', as the default instantiation). Valent is a super category dominating the subdivisions of subject, object and complement. For Japanese, in addition to the distinction between valent and adjunct, two primitive grammatical relations appear to be necessitated: topic and extraposee - for the examples of extraposition, refer to (4.51).

The overall system of grammatical functions for Japanese is manifested in Figure (4.52). The diagram is in the shape of an inheritance hierarchy where dependent is by default a pre-dependent, hence it is unspecified. Extraposee is the exception in the figure, which overrides the default property of pre-dependent.
There are plenty of good reasons for distinguishing subject and object in elucidating such phenomena as quantifier floating (Haig 1980; Kuroda 1980, 1983; Miyagawa 1988, 1989) and subject honorification (Harada 1976a; Shibatani 1977, 1978) amongst others. To take the latter case as example, the verbal complex consisting of the honorific prefix 'o-', the adverbial form of a verb and the particle 'ni' followed by the other full verb 'naru' (to become) is utilised in expressing honour for the subject. In (4.53) the first sentence is acceptable whose verbal sequence 'o-tashiname ni naru' (to reprove) shows respect to the subject 'taisa' (colonel), whereas the second one is pragmatically inadmissible where the honorific verb fails to honorify 'taisa' since it is complemented by the object marker 'o'. This fact therefore supports the existence of subjecthood in Japanese.

(4.53)
Taisa wa souchou o o-tashiname ni naru.
colonel sergeant to reprove to become  
(Colonel reproves Sergeant.)

*Taisa o souchou wa o-tashiname ni naru.
colonel sergeant to reprove to become  
(Sergeant reproves Colonel.)

Further, the dependency type 'topic' must be recognised to account for sentences like (4.54) where the first nominal item 'itaria' (Italy) should be treated as the topic in the sentence, given that the second noun 'koohii' (coffee) is the subject.

(4.54)
Itaria wa/ga koohii wa/ga oishii.
Italy coffee tasty.  
(As for Italy, coffee is tasty.)

4.4.3 Inflections

Finally, the morphosyntactic inflections for Japanese are outlined. The Japanese noun, in contrast to English, has no inflections. Although there are a few suffixes that signal the plurality of pronouns and some common nouns, they are not qualified as syntactic inflections owing to their
irrelevance to syntax. For instance, the plural suffix ‘-ra’ and ‘-tachi’ are optionally attached to pronouns and regular nouns signifying humans, animals, flowers, stones and so forth. There are however no agreement rules that force us to recognise the number feature; as a result, all sentences in (4.55) are grammatical. In consequence, these plural suffixes for Japanese nouns ought to be considered in the domain of derivational morphology.

(4.55) kono tori wa natsu ni watari o okonau.
this bird summer in migration do
(This bird does its migration in summer.)

kono toritachi wa natsu ni watari o okonau.
this birds summer in migration do
(These birds do their migration in summer.)

korerano tori wa natsu ni watari o okonau.
these bird summer in migration do
(These birds do their migration in summer.)

korerano toritachi wa natsu ni watari o okonau.
these birds summer in migration do
(These birds do their migration in summer.)

Inflections yield distinct word-classes that cut across the distinctions between word-classes in Japanese. The three inflectional categories, i.e. full verb, auxiliary verb and adjective share the same inflections. It is because of this similarity that I assume that these three groups are subsumed within the super-class ‘verb’. The inflectional values are relevant to syntax in the sense that they are selected by parent verbs or particles. Hence, seven inflections, which Table 4.1 has already introduced, will be presumed for inflectional lexemes belonging to the word-class ‘verb’. Table 4.2 below shows all inflectional values and some examples of the inflections of ‘full verb’, ‘auxiliary verb’ and ‘adjective’.

5 An alternative inflectional form for adverbial to that shown in Table 4.2 is used, followed by the auxiliary verb ‘ta’ (past) or the particle ‘te’. This alternative however, merely results from a phonological process, so it will be unnecessary to recognise it as a distinctive inflectional form. For the inflections of ‘ta’ (past), refer to Table 4.2.
Table 4.2 Inflections of ‘full verb’, ‘auxiliary verb’ and ‘adjective’

<table>
<thead>
<tr>
<th>Inflections</th>
<th>full verb (regular)</th>
<th>full verb (irregular)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>type I</td>
<td>type II</td>
</tr>
<tr>
<td>to cut</td>
<td>ki</td>
<td>kake</td>
</tr>
<tr>
<td>to wear</td>
<td>ki</td>
<td>kake</td>
</tr>
<tr>
<td>to hang</td>
<td>ki</td>
<td>kake</td>
</tr>
<tr>
<td>to do</td>
<td>kake</td>
<td>su-ru</td>
</tr>
<tr>
<td>to come</td>
<td>shi</td>
<td>kake-ru</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Inflections</th>
<th>auxiliary verb</th>
<th>adjective</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>causative</td>
<td>copula</td>
</tr>
<tr>
<td>Irrealis</td>
<td>se</td>
<td>-</td>
</tr>
<tr>
<td>Adverbial</td>
<td>se</td>
<td>de</td>
</tr>
<tr>
<td>Conclusive</td>
<td>se-ru</td>
<td>da</td>
</tr>
<tr>
<td>Attributive</td>
<td>se-ru</td>
<td>na/taru</td>
</tr>
<tr>
<td>Hypothetical</td>
<td>se-re</td>
<td>na-ra</td>
</tr>
<tr>
<td>Imperative</td>
<td>se-ro/yo</td>
<td>tare</td>
</tr>
<tr>
<td>Cohortative</td>
<td>se-yo</td>
<td>da-ro</td>
</tr>
</tbody>
</table>

N.B. - indicates a segmental boundary between a stem and an inflectional suffix.

### 4.4.4 Notation

Thus far, the Word Grammar word-classes, grammatical functions, and inflections for Japanese have been summarised. They will be employed for all subsequent dependency analyses for Japanese data. Table 4.3 is the summary of notation to be used.
Table 4.3
Notation for the Word Grammar word-classes, grammatical functions, and inflections for Japanese

<table>
<thead>
<tr>
<th>Word-classes</th>
<th>Inflections</th>
<th>Grammatical Functions</th>
</tr>
</thead>
<tbody>
<tr>
<td>full verb</td>
<td>V</td>
<td>irrealis</td>
</tr>
<tr>
<td>auxiliary verb</td>
<td>v</td>
<td>adverbial</td>
</tr>
<tr>
<td>adjective</td>
<td>J</td>
<td>conclusive</td>
</tr>
<tr>
<td>pronoun</td>
<td>n</td>
<td>attributive</td>
</tr>
<tr>
<td>common noun</td>
<td>N</td>
<td>hypothetical</td>
</tr>
<tr>
<td>proper noun</td>
<td>nN</td>
<td>imperative</td>
</tr>
<tr>
<td>adjectival noun</td>
<td>jN</td>
<td>cohortive</td>
</tr>
<tr>
<td>verbal noun</td>
<td>vN</td>
<td></td>
</tr>
<tr>
<td>classifier</td>
<td>qN</td>
<td></td>
</tr>
<tr>
<td>digit</td>
<td>dN</td>
<td></td>
</tr>
<tr>
<td>magnitude indicator</td>
<td>mN</td>
<td></td>
</tr>
<tr>
<td>function particle</td>
<td>fP</td>
<td></td>
</tr>
<tr>
<td>postposition particle</td>
<td>pP</td>
<td></td>
</tr>
<tr>
<td>discourse particle</td>
<td>dP</td>
<td></td>
</tr>
<tr>
<td>coordination particle</td>
<td>cP</td>
<td></td>
</tr>
<tr>
<td>adverb</td>
<td>A</td>
<td></td>
</tr>
</tbody>
</table>

4.5 The minimal processing unit

In the preceding sections, I suggested what should constitute syntactic wordhood in Japanese. As far as syntax concerned, this is important in applying the most plausible dependency analyses and as the basis for making generalisations in the grammar. It is therefore a syntactic word that could count as the minimal syntactic processing unit.

At the very initial stage of the comprehension process, that is, the word recognition stages, it is uncontroversial that the word recognition point normally occurs before the listeners have heard the entire word (Harley 1995). This means that as we hear speech or read text, the recognition process - in Word Grammar terms this operation would be comparable to the identification of temporary words - is instantaneously taking place. Bearing in mind that word recognition is the onset of the syntactic process,
we can assume that syntactic parsing should also be an on-line process unlike something like the "sausage machine" proposed by Frazier and Fodor (1978) in which packets of about six input words are syntactically structured by turns. Marslen-Wilson (1973, 1975, 1976) supports this view with shadowing experiments which show that both syntactic and semantic processing are set off virtually immediately and are not delayed until a syntactic unit (whole phrase) has been heard. It is therefore reasonable to conclude that processing defines its own units, so there is no need to divide the input in advance into separate processing units.

On the other hand, it is important for our purposes to be able to measure the input in terms of some kind of minimum processing unit which is relevant to syntactic difficulty. What we need is some way of measuring the 'length' of an utterance, prior to measuring its syntactic difficulty, which does not depend too much on analytical assumptions about syntactic words.

Let us take the string of words 'boku wa' for instance. If we adopt the lexicalist approach, this string is syntactically one word with '-wa' as a derivational suffix - cf. (4.24). Hence, before having lexical access, a morphological analysis must divide 'boku wa' into 'boku' and 'wa'. This may be the correct analysis, considering the fact that each past form of an English regular verb is not stored in our mental grammar, but is generated by applying by default the morphological rule to the verb (Marcus, Pinker, Ullman, Hollander, Rosen & Xu 1992; Pinker 1994). On the other hand, I have argued above that 'wa' is actually a separate syntactic word. The pivotal difference between the lexicalist and the Word Grammar approaches to 'boku wa' lies in the stages at which its lexical or syntactic analysis materialises. The figure (4.56) below compares the two analyses and their effects on processing. The salient point for computing the syntactic difficulty of 'boku wa' is to provide the same amount of processing load for it irrespective of which syntactic analysis one may apply, lexicalist or Word Grammar.
(4.56)

<the lexicalist model>  <the Word Grammar model>

**lexical access**  **lexical access**
and  and
morphological  syntactic
analysis  analysis

```
  BOKU  BOKU
     w₁ = x₁
   w₁ → wa
     w₂

  word recognition  word recognition
  w₁  w₁  w₂
```

Input: /bokuwa/  Input: /bokuwa/

The processing load \( x₁^o \) = the processing loads \( x₁ + x₂ \)

To summarise this discussion of word-hood in Japanese, therefore, I shall assume a 'syntactic' rather than a morphological treatment for all the morphemes discussed except for those which are clearly inflection markers on verbs.

The experiment reported in Chapter 5 tentatively supports the Word-Grammar analysis, in which each morpheme contributes to the difficulty of processing, rather than one in which morphologically complex words count no more than morphologically simple words. On the other hand, it is also clear that morphemes do not all contribute the same amount of difficulty. For example, idioms and cliches presumably contribute less per word than freely combined words (Ellis 1996). However, this kind of variation is not relevant to the experiments reported here because the experimental material deliberately excluded fixed phrases of all kinds. Moreover, the experiment mentioned above seems to show that functional morphemes such as auxiliary verbs and particles and content morph contribute differently.
5 Variations in memory load for different kinds of words

As explained in Chapter 1, Hudson (1995a) introduced a paradigm to test short-term memory retention, where subjects are instructed to write down sentences as soon as they have been read out. This experimental technique has a history in this type of research such as the elicited imitation tasks of Miller and Isard (1964), Lado (1965), Epstein (1967), Cook (1979) and Bley-Vroman and Chaudron (1994). In Miller and Isard's 'immediate recall' experiment, subjects were instructed to memorise six tape-recorded 22-word sentences with various degrees of self-embedding. The experiment I here report utilises the same technique to explore memory load on different strings of words. The idea is that the greater the memory load of these items, the less well they will be recalled.

5.1 Experiment 1

This experiment was prompted by the assumption that memory cost varies for different types of word-strings; a slightly more elaborate prediction would be that given equal phonological length, a combination of a content word and a function word might involve greater memory cost than a single content word.

The main aim of the experiment was to investigate the relative memorability of these strings of words. The content words are, so far as this experiment is concerned, any nouns including adjectival nouns and verbal nouns, adjectives, and full verbs. On the other hand, function words are particles and auxiliary verbs. This experiment was not designed to estimate the respective memory load of different word-classes among the content or function words. This investigation was a tentative pilot experiment, which is worth reporting because it shows clearly that grammatical structure affects short-term memory independently of phonology.

In short-term memory tasks, there are many determinants of memorability, e.g.
• chunking effects (Miller 1956, 1958).
• short-term memory span affected by the acoustic confusability of the items (Conrad 1964).
• word length in terms of syllables and pronunciation time (Baddeley &
Hitch 1974).

- word frequency and familiarity effects such as those of long-term lexical knowledge (Brown & Hulme 1992).
- effects of long-term semantic knowledge on short-term memory for word strings (Cook 1979).
- syntactic influences (Epstein 1967).

As will be seen from the following section, memorability might be strongly affected by the cohesion of such neighbouring words as adjective and noun strings, which were not properly controlled in this experimental design. In addition, there is only a small sample of items in the experiment. It is therefore possible that the word-classes as represented in the stimuli are confounded with other factors such as familiarity, pronounciability and word-length.

5.2 Materials and Procedure

Materials

There were thirteen sequences in all, which are listed in (5.1) below. Three of them were natural Japanese sentences, while the rest were random strings of words. All the sequences were composed of familiar Japanese words. The phonological length of the sequences was constant in terms of the number of syllables or morae; all the strings consisted of either 48 or 60 morae.\(^1\) The consideration of morae was relevant to the experiment, because the number of morae approximately corresponded to the number of Japanese hiragana or katakana letters which the subjects used to write the sequences down. Every string contained twelve meaningful units; one unit consisted of four morae for 48-mora strings, and of five morae for 60-mora strings. In each unit, there was one content word which might be accompanied by one or more than one function words. Table 5.1 below shows the characteristics of the sequences used in the experiment, which are listed in order of

\(^1\) From a series of mini-scale pilot tests, conducted before Experiment 1, which I had conducted with several native Japanese undergraduate students at University College London, I had learned that a random sequence of eleven meaningful units with the length of about 45 morae was the limit for pure rote recall.

142
presentation. ‘µ’ stands for the number of morae. In the sequence labels, ‘dic’ stands for ‘dictionary’, i.e. content words as opposed to function words, and ‘scrb’ means ‘scrambled’.

**Table 5.1 Characteristics of the sequences used in Experiment 1**

<table>
<thead>
<tr>
<th>order</th>
<th>label</th>
<th>µ</th>
<th>characteristics of 12 units</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>nat sent 1</td>
<td>48</td>
<td>grammatically structured natural sentence.</td>
</tr>
<tr>
<td>2</td>
<td>nat sent 2</td>
<td>48</td>
<td>grammatically structured natural sentence.</td>
</tr>
<tr>
<td>3</td>
<td>shrt dic 1</td>
<td>48</td>
<td>4N, 4V, 4J, arranged randomly.</td>
</tr>
<tr>
<td>4</td>
<td>shrt d+p 2</td>
<td>60</td>
<td>4(N+p), 4(V+p), 4(J+p), arranged at random.</td>
</tr>
<tr>
<td>5</td>
<td>nmb+clsf 48</td>
<td>48</td>
<td>12(No+c), arranged at random.</td>
</tr>
<tr>
<td>6</td>
<td>auxlries 1</td>
<td>48</td>
<td>4(N+v), 4(V+v), 4(JN+v), arranged at random.</td>
</tr>
<tr>
<td>7</td>
<td>long dic 1</td>
<td>60</td>
<td>4N, 4V, 4J, arranged at random.</td>
</tr>
<tr>
<td>8</td>
<td>nat sent 3</td>
<td>48</td>
<td>grammatically structured natural sentence.</td>
</tr>
<tr>
<td>9</td>
<td>auxlries 2</td>
<td>48</td>
<td>4(N+v), 4(V+v), 4(JN+v), arranged at random.</td>
</tr>
<tr>
<td>10</td>
<td>long dic 2</td>
<td>60</td>
<td>4N, 4V, 4J, arranged at random.</td>
</tr>
<tr>
<td>11</td>
<td>scrb nat 2</td>
<td>48</td>
<td>12 units of ‘nat sent 2’, arranged at random.</td>
</tr>
<tr>
<td>12</td>
<td>shrt dic 2</td>
<td>48</td>
<td>4N, 4V, 4J, arranged at random.</td>
</tr>
<tr>
<td>13</td>
<td>shrt d+p 1</td>
<td>60</td>
<td>4(N+p), 4(V+p), 4(J+p), arranged at random.</td>
</tr>
</tbody>
</table>

KEY: N = noun; V = full verb; J = adjective; p = particle; No = numeral; c = classifier; v = one or two auxiliary verbs.

(5.1) **Sequences used for Experiment 1**

N.B. I indicates a boundary between meaningful units in a sequence.

1. nat sent 1

Akiya de wal₁ at te mo,₁ tanin no₁ ie no₄
unoccupied house topic to be although stranger of house of
N pP fP V:a cP pP N pP N pP

shikichi nil₁₅ katte nil₆ ashi ol₁ humiirerul₁₆ koto wa,l₁₉ houritsu
premises on own way foot object to step into thing subject law
N pP jN pP N fP V:t N fP N

nil₁₁₀ hanshi tal₁₁ koui dal₁₁₂
to to be against past act copula
pP V:a v:t N v:c

(Stepping into the premises of a stranger's house without permission, even though it is an unoccupied house, is an illegal act.)
2. nat sent 2

Boku wa sofa nil koshi ol, oroshi tele asa no, nokori no.

I subject sofa in waist object to lower morning of rest of

koohii ol, nomi nagara atama no, naka ol sokushii

coffee object to drink while head no inside object a little

seirishi-te-mi ta, to try clearing past

(Sitting in the sofa and drinking the remaining morning coffee, I tried clearing my head off a little.)

3. shrt dic 1

huusen hurl, muzukashi taberu, akarui, depaato

balloon old difficult to eat bright department store

hataora to ask tasty mouse to arrange gift to be born

4. shrt d+p 2

atama gal komakaku tel kurai tol kangae tele hataraake bal

head subject fine dark to think to work if

nuigurumi tol, hohoemu tol, tsutsumashiku tele itaria no

stuffed toy object to smile modest Italy of

gyuniku wal, ureshikere bal mayot tel

meat topic glad if to wonder

5. nmb+clsf

jup pon hachi mail juu-ichi jil, ni-jut tou

10 bottles 8 thin flat objects 11 o'clock 20 animals

ichi il, ni wa san kol, go-juu-san pun, kyyu da, 1 place 2 birds 3 pieces 53 minutes 9 machines

144
6. auxlries 1

kirei dal\textsubscript{1} hon desul\textsubscript{2} mi mashi tal\textsubscript{4} uso dal\textsubscript{4} shizuka
beautiful copula book copula to see polite past lie copula quiet

dal\textsubscript{5} kaeri masul\textsubscript{6} ason dal\textsubscript{7} yuikai nal\textsubscript{8} namida dal\textsubscript{9}
copula to return polite to play past pleasant copula tear copula

genki nal\textsubscript{10} hana desul\textsubscript{11} kikoe tal\textsubscript{12}
healthy copula flower copula to sound past
jN v:t N v:c V:a v:c

7. long dic 1

reizoukol\textsubscript{5} konpyuutaal\textsubscript{2} chirakasul\textsubscript{3} kagayakashiil\textsubscript{4} tamagol\textsubscript{5}
refrigerator computer to scatter glorious egg
N N v:c J:c N

itametsukerul\textsubscript{6} atsukamashiil\textsubscript{7} moushikomul\textsubscript{8} suzushiil\textsubscript{9} yakusokul\textsubscript{10}
to torment presumptuous to apply cool promise
V:c J:c V:c J:c N

kashikoil\textsubscript{11} wakimaerul\textsubscript{12}
wise to distinguish
J:c V:c

8. nat sent 3

Tonari nol\textsubscript{1} musume wa,l\textsubscript{2} boku nol\textsubscript{3} tekubi nol\textsubscript{4} kusuriyubi ol\textsubscript{5}
next of girl subject 1 of wrist on little finger object
N pP N fP n pP N pP N fP

oi te,l\textsubscript{6} katachi nol\textsubscript{7} sadamara nail\textsubscript{8} kimyou nal\textsubscript{9} zukai ol\textsubscript{10}
to put shape of to be fixed negative strange copula figure object
V:a cP N pP V:i v:t jN v:t N fP

soko nol\textsubscript{11} kai ta,l\textsubscript{12}
there on to draw past
n pP V:a v:c

(The girl next to me put her little finger on my wrist and drew a shapeless and strange figure there.)
9. auxlries 2

nigiyaka dal1 hune desul2 yomi mashi tal3 kome dal4 kai
bustling copula ship copula to read polite past rice copula to buy

masul5 hima nal6 usagi desul7 benri nal8 ton dal9
polite leisure copula rabbit copula convenient copula to fly past

tomat tal10 kirai dal11 anata dal12
to stop past to dislike copula you copula
V:a v:c jN v:c n v:c

10. long dic 2

nagusamerul1 mayoneezul2 isogashi3 dairisekil4 urayamashi5 otoroerul6
to comfort mayonnaise busy marble envious to decline

shiraberul7 kanashi8 keisatsukan9 omoidasul10 niwatoril11
to examine sad police officer to remember fowl
V:c J:c N V:c N

misuborashiil12
shabby
J:c

11. scrb nat 2

nokori no1 sofaii2 koshi ol3 asa no4 Boku wai5 nomi
rest of sofa in waist object morning of I subject to drink
N pP N pP N fP N pP n fP V:a

nagaral6 seirishi-te-mi tal7 sukoshil8 kohii ol9 atama no10
while to try clearing past a little coffee object head of
A cP V:a cP N pP N fP V:a

oroshi tel11 naka ol12
to lower inside object
V:a cP N pP

12. shrt dic 2

hatarakul1 kurail2 atamal3 hohoemul4 tsutsumashiil5 kangaerul6 itarial7
to work dark head to smile modest to think Italy
V:c J:c N V:c J:c V:c nN

ureshiil8 gyuunikul9 komakail10 mayoul11 nuigurumil12
glad meat fine to wonder stuffed toy
J:c N J:c V:c N
13. shrt d+p 1

hurukere bal₁ nezumi gal₂ totonoe tel₃ akaruku tel₄ huusen walg old if mouse subject to arrange bright balloon topic
tabere bal₆ umare tel₇ depaato nil₈ muzukashiku tel₉
to eat if to be born department store to difficult
V:₃ pP V:₃ cP N pP J:₄ cP
okurimono o₁₀ tazuneru tol₁₁ oishikere bal₁₂
gift object to ask tasty if
N fP V:c cP J:₃ pP

The following are notes on the thirteen sequences:

• ‘nat sent 1’, ‘nat sent 2’ and ‘nat sent 3’ were the three natural Japanese sentences, while ‘scrb nat 2’ was the scrambled version of ‘nat sent 2’, in which the twelve units of ‘nat sent 2’ were arranged in a random order.

• As regards the two pairs, ‘shrt die 1’ and ‘shrt die d+p 1’, and ‘shrt die 2’ and ‘shrt die d+p 2’, each pair contained exactly the same content words; of course most of the verbs and adjectives were inflected. The sole disparity within each pair was that the ‘d+p’ versions accommodated one particle in every unit. Hence, these strings were longer by twelve morae in comparison to their particle-less counterparts.

• ‘long die 1’ and ‘long die 2’ had entirely different content words from those of ‘shrt die 1’ and ‘shrt die 2’. On the other hand, their lengths were equal to those of ‘shrt d+p 1’ and ‘shrt d+p 2’. For the verbs and adjectives of the four ‘die’ strings, regardless of length, only the citation form was used.

• There were two strings containing auxiliary verbs, namely ‘auxlries 1’ and ‘auxlries 2’. Both sequences held thirteen auxiliaries; each unit contained one auxiliary verb except the third units of both sequences which contained two auxiliaries, viz. mashi (polite) and ta (past).

• ‘nmb+clsf’ was a random sequence of twelve numeral quantifiers, each consisting of a numeral and a classifier. The numerals consisted of one or two digits which might be accompanied by a magnitude indicator.

• Although random sequences were intended to preclude syntactic analysis, some random sequences turned out to contain a string of two units which allowed an interpretation where the preceding unit modified the succeeding unit. A common pattern was a string of two units consisting of an adjective followed by a noun. This type of string may have been analysed
as one chunk, which might encourage a better performance than had been expected on the units forming the chunk. My intention was to avoid chunks, but to a certain extent I failed to achieve it in designing the sequences.

**Procedures**

Experiment 1 was conducted in the summer of 1996. My subjects were one hundred native Japanese overseas students, aged on average 20, at the Institute of International Education in London. In each test, I read the list of sequences shown above in (5.1) in that order to a group of between 10 and 20 subjects. The average rate of presentation was 0.15 seconds per mora. For the three natural sentences, each sentence was grouped into three phrases whose boundaries are marked by a comma in (5.1) and approximately one mora of pause was made after each phrase. As for the rest of the sequences, each unit formed one group and about one mora of pause was placed between any adjacent units; hence there were eleven pauses for each random sequences. All the sequences were read in a flat fashion - no particular intonation was overlaid - with standard Japanese pitch accent. My instructions to the subjects were that they might write each sequence down as soon as I had read it out, and stressed only accuracy of recall. The subjects were allowed to write in any Japanese letters: two alphabets *hiragana* and *katakana* as well as Chinese characters *kanji*. Also Arabic numerals were permitted to be used for digits. The sequences were read only once and adequate time was given after the presentation of each sequence in order for all the subjects to complete their reproduction of the sequence. Thus the experiment examined the subjects’ short-term retention.

The scoring system was based on the twelve units in a string; each unit was scored either as a success or as a failure. Since each sequence had twelve units, the sum of the successes and failures for each string should always be twelve. A recall of any unit was a success only if it was a verbatim copy of the original or discrepancies could be attributed to mishearing. Any other response including synonyms and abbreviations of words was counted as a fail; subjects were told to write exact words. Concerning displaced units, if disarrangement was owing to missing units, only these were counted as failures, and if displacement was caused by a reversal of units, scores were calculated in such a way that the reversed units could obtain as many successes as possible.
5.3 Results

First of all, it can be seen in Table 5.2 that the subjects attained an average of 62.5% units correct. Also as can be seen from the table, subjects' performance on the three natural sentences, an average of 81.55% correct, was better than that on the other random sequences, an average of 56.79% correct, which suggests that syntactic structure does affect short-term retention.

Table 5.3 with its accompanying graph for average values below shows the number of failures for each unit in all sequences and its serial-position curve for the experiment.

Table 5.2 The number of Failures and Successes for each sequence

<table>
<thead>
<tr>
<th></th>
<th>Failures</th>
<th>Successes</th>
<th>% of Successes</th>
</tr>
</thead>
<tbody>
<tr>
<td>nat sent 1</td>
<td>306</td>
<td>894</td>
<td>74.50</td>
</tr>
<tr>
<td>nat sent 2</td>
<td>130</td>
<td>1070</td>
<td>89.16</td>
</tr>
<tr>
<td>nat sent 3</td>
<td>228</td>
<td>972</td>
<td>81.00</td>
</tr>
<tr>
<td>subtotal</td>
<td>664</td>
<td>2936</td>
<td>81.55</td>
</tr>
<tr>
<td>shrt dic 1</td>
<td>577</td>
<td>623</td>
<td>51.91</td>
</tr>
<tr>
<td>shrt dic 2</td>
<td>422</td>
<td>778</td>
<td>64.83</td>
</tr>
<tr>
<td>auxlries 1</td>
<td>474</td>
<td>726</td>
<td>60.50</td>
</tr>
<tr>
<td>auxlries 2</td>
<td>614</td>
<td>586</td>
<td>48.83</td>
</tr>
<tr>
<td>nmb+clsf</td>
<td>376</td>
<td>824</td>
<td>68.66</td>
</tr>
<tr>
<td>scrb nat 2</td>
<td>494</td>
<td>706</td>
<td>58.83</td>
</tr>
<tr>
<td>long dic 1</td>
<td>519</td>
<td>681</td>
<td>56.75</td>
</tr>
<tr>
<td>long dic 2</td>
<td>521</td>
<td>679</td>
<td>56.58</td>
</tr>
<tr>
<td>shrt d+p 1</td>
<td>613</td>
<td>587</td>
<td>48.91</td>
</tr>
<tr>
<td>shrt d+p 2</td>
<td>575</td>
<td>625</td>
<td>52.08</td>
</tr>
<tr>
<td>subtotal</td>
<td>5185</td>
<td>6815</td>
<td>56.79</td>
</tr>
<tr>
<td>total</td>
<td>5849</td>
<td>9751</td>
<td>62.50</td>
</tr>
<tr>
<td>average</td>
<td>450</td>
<td>750</td>
<td>62.50</td>
</tr>
</tbody>
</table>
Table 5.3 Number of failures for each unit in all sequences used in Experiment 1

<table>
<thead>
<tr>
<th>unit number</th>
<th>u1</th>
<th>u2</th>
<th>u3</th>
<th>u4</th>
<th>u5</th>
<th>u6</th>
<th>u7</th>
<th>u8</th>
<th>u9</th>
<th>u10</th>
<th>u11</th>
<th>u12</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>nat sent 1</td>
<td>0</td>
<td>0</td>
<td>8</td>
<td>41</td>
<td>43</td>
<td>43</td>
<td>35</td>
<td>37</td>
<td>15</td>
<td>13</td>
<td>29</td>
<td>42</td>
<td>306</td>
</tr>
<tr>
<td>nat sent 2</td>
<td>0</td>
<td>0</td>
<td>8</td>
<td>10</td>
<td>30</td>
<td>45</td>
<td>4</td>
<td>6</td>
<td>1</td>
<td>4</td>
<td>19</td>
<td>3</td>
<td>130</td>
</tr>
<tr>
<td>nat sent 3</td>
<td>1</td>
<td>1</td>
<td>12</td>
<td>44</td>
<td>46</td>
<td>26</td>
<td>13</td>
<td>17</td>
<td>11</td>
<td>23</td>
<td>29</td>
<td>5</td>
<td>228</td>
</tr>
<tr>
<td>shrt dic 1</td>
<td>1</td>
<td>2</td>
<td>42</td>
<td>67</td>
<td>68</td>
<td>47</td>
<td>90</td>
<td>55</td>
<td>46</td>
<td>86</td>
<td>32</td>
<td>41</td>
<td>577</td>
</tr>
<tr>
<td>shrt dic 2</td>
<td>1</td>
<td>7</td>
<td>11</td>
<td>68</td>
<td>63</td>
<td>46</td>
<td>28</td>
<td>57</td>
<td>36</td>
<td>78</td>
<td>13</td>
<td>14</td>
<td>422</td>
</tr>
<tr>
<td>auxlries 1</td>
<td>4</td>
<td>11</td>
<td>19</td>
<td>49</td>
<td>58</td>
<td>65</td>
<td>65</td>
<td>40</td>
<td>55</td>
<td>56</td>
<td>44</td>
<td>8</td>
<td>474</td>
</tr>
<tr>
<td>auxlries 2</td>
<td>14</td>
<td>25</td>
<td>67</td>
<td>71</td>
<td>74</td>
<td>56</td>
<td>61</td>
<td>87</td>
<td>67</td>
<td>50</td>
<td>34</td>
<td>8</td>
<td>614</td>
</tr>
<tr>
<td>nmb+clsf</td>
<td>2</td>
<td>7</td>
<td>27</td>
<td>47</td>
<td>67</td>
<td>49</td>
<td>54</td>
<td>38</td>
<td>32</td>
<td>24</td>
<td>24</td>
<td>5</td>
<td>376</td>
</tr>
<tr>
<td>scrb nat 2</td>
<td>0</td>
<td>10</td>
<td>25</td>
<td>36</td>
<td>48</td>
<td>82</td>
<td>60</td>
<td>65</td>
<td>47</td>
<td>51</td>
<td>43</td>
<td>27</td>
<td>494</td>
</tr>
<tr>
<td>long dic 1</td>
<td>1</td>
<td>4</td>
<td>33</td>
<td>59</td>
<td>61</td>
<td>83</td>
<td>70</td>
<td>72</td>
<td>46</td>
<td>56</td>
<td>29</td>
<td>5</td>
<td>519</td>
</tr>
<tr>
<td>long dic 2</td>
<td>3</td>
<td>2</td>
<td>52</td>
<td>46</td>
<td>76</td>
<td>76</td>
<td>87</td>
<td>29</td>
<td>30</td>
<td>70</td>
<td>43</td>
<td>7</td>
<td>521</td>
</tr>
<tr>
<td>shrt d+p 1</td>
<td>2</td>
<td>14</td>
<td>27</td>
<td>69</td>
<td>82</td>
<td>83</td>
<td>64</td>
<td>68</td>
<td>49</td>
<td>56</td>
<td>16</td>
<td>1</td>
<td>613</td>
</tr>
<tr>
<td>shrt d+p 2</td>
<td>4</td>
<td>11</td>
<td>43</td>
<td>59</td>
<td>83</td>
<td>85</td>
<td>95</td>
<td>86</td>
<td>9</td>
<td>15</td>
<td>59</td>
<td>26</td>
<td>575</td>
</tr>
<tr>
<td>total</td>
<td>33</td>
<td>94</td>
<td>374</td>
<td>666</td>
<td>799</td>
<td>786</td>
<td>745</td>
<td>653</td>
<td>575</td>
<td>454</td>
<td>207</td>
<td>5849</td>
<td></td>
</tr>
</tbody>
</table>

Average

N.B. 'average' displays an average value for each unit of all thirteen sequences.

Graph 5.1 Average value for failures for each unit in all sequences

Since the thirteen sequences differ in their phonological lengths and in whether they were natural sentences or random sequences, they are first grouped into the following three categories:
[Category I] 48-mora natural sentences
'nat sent 1', 'nat sent 2' and 'nat sent 3'

[Category II] 48-mora random sequences
'shrt dic 1', 'shrt dic 2', 'auxlries 1', 'auxlries 2', 'nmb+clsf'
and 'scrb nat 2'

[Category III] 60-mora random sequences
'long dic 1' and 'long dic 2', 'shrt d+p 1' and 'shrt d+p 2'

Then, the raw data comprising the number of failures for these three categories is processed by a one-way analysis of variance of a repeated-measures design, since a hundred subjects all completed tests in which they reproduced all the thirteen sequences.

Table 5.4 The number of Failures and Successes for Category I

<table>
<thead>
<tr>
<th></th>
<th>Failures</th>
<th>Successes</th>
<th>% of Successes</th>
</tr>
</thead>
<tbody>
<tr>
<td>nat sent 1</td>
<td>306</td>
<td>894</td>
<td>74.50</td>
</tr>
<tr>
<td>nat sent 2</td>
<td>130</td>
<td>1070</td>
<td>89.16</td>
</tr>
<tr>
<td>nat sent 3</td>
<td>228</td>
<td>972</td>
<td>81.00</td>
</tr>
</tbody>
</table>

The results of the analysis of variance applied to the data in Table 5.4 shows that there is a significant difference amongst the three natural sentences, $F(2,198) = 37.38, p < .05$. The difference in results for these three sentences may arise from their dependency distances, which will be explained in the next section.

Table 5.5 The number of Failures and Successes for Category II

<table>
<thead>
<tr>
<th></th>
<th>Failures</th>
<th>Successes</th>
<th>% of Successes</th>
</tr>
</thead>
<tbody>
<tr>
<td>shrt dic 1</td>
<td>577</td>
<td>623</td>
<td>51.91</td>
</tr>
<tr>
<td>shrt dic 2</td>
<td>422</td>
<td>778</td>
<td>64.83</td>
</tr>
<tr>
<td>auxlries 1</td>
<td>474</td>
<td>726</td>
<td>60.50</td>
</tr>
<tr>
<td>auxlries 2</td>
<td>614</td>
<td>586</td>
<td>48.83</td>
</tr>
<tr>
<td>nmb+clsf</td>
<td>376</td>
<td>824</td>
<td>68.66</td>
</tr>
<tr>
<td>scrb nat 2</td>
<td>494</td>
<td>706</td>
<td>58.83</td>
</tr>
</tbody>
</table>
Next, analysis of variance comparisons amongst Category II, 48-mora random sequences, shown in Table 5.5 reveal that there were significant differences across sequences, $F(5,495) = 46.88$, $p < .05$. Considering the fact that this category contains different types of sequence, accommodating four types, these statistic differences were expected. However, it is hard to distinguish the effects of the four "types" from those of individual sequences on the basis of so few examples of each type.

Table 5.6 The number of Failures and Successes for Category III

<table>
<thead>
<tr>
<th></th>
<th>Failures</th>
<th>Successes</th>
<th>% of Successes</th>
</tr>
</thead>
<tbody>
<tr>
<td>long dic 1</td>
<td>519</td>
<td>681</td>
<td>56.75</td>
</tr>
<tr>
<td>long dic 2</td>
<td>521</td>
<td>679</td>
<td>56.58</td>
</tr>
<tr>
<td>shrt d+p 1</td>
<td>613</td>
<td>587</td>
<td>48.91</td>
</tr>
<tr>
<td>shrt d+p 2</td>
<td>575</td>
<td>625</td>
<td>52.08</td>
</tr>
</tbody>
</table>

Further, as we found in Category II, the results of analysis of variance on Category III, whose data are displayed in Table 5.6, also show that there were differences amongst 60-morae random sequences in Categories III, $F(3,297) = 13.59$, $p < .05$. These differences were expected as well because Category III contains two morphologically different types of sequences. However, in this case the two general types are clearly more homogeneous than in Category II.

Bearing in mind that the prime objective of the experiment was to explore the relative memorability of the strings of words consisting of content words only and of content plus function words, we will proceed to make further comparisons between pairs of means in Categories II and III. For these comparisons, the procedure often referred to as the protected t will be used, since these two categories meet the requirement that the overall F for an analysis of variance is significant. In view of the constant phonological length of the sequences in each category, the most crucial comparisons should be between the monomorphemic content-word sequences 'shrt dic 1+2' and the bi-morphemic content plus function-word sequences 'auxrries 1+2' in Category II and between 'long dic 1+2' and 'shrt d+p 1+2', which are the 60 mora-length equivalents in Category III. Table 5.7 displays the data for these four groups and the results on these two comparisons are presented.
in Tables 5.8 and 5.9 below.

Table 5.7 The data for 'shrt die 1+2', 'auxlries 1+2' in Category II and 'long die 1+2' and 'shrt d+p 1+2' in Category III

<table>
<thead>
<tr>
<th></th>
<th>Failures</th>
<th>Successes</th>
<th>% of Successes</th>
</tr>
</thead>
<tbody>
<tr>
<td>shrt die 1+2</td>
<td>999</td>
<td>1401</td>
<td>58.37</td>
</tr>
<tr>
<td>auxlries 1+2</td>
<td>1088</td>
<td>1312</td>
<td>54.66</td>
</tr>
<tr>
<td>long die 1+2</td>
<td>1040</td>
<td>1360</td>
<td>56.66</td>
</tr>
<tr>
<td>shrt d+p 1+2</td>
<td>1188</td>
<td>1212</td>
<td>50.50</td>
</tr>
</tbody>
</table>

Table 5.8 Protected $t$ between 'shrt die 1+2' and 'auxlries 1+2'

<table>
<thead>
<tr>
<th></th>
<th>shrt die 1+2</th>
<th>auxlries 1+2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\bar{X}_1$</td>
<td>4.99</td>
<td>5.44</td>
</tr>
<tr>
<td>$n_1$</td>
<td>200</td>
<td>200</td>
</tr>
<tr>
<td>$\text{MS}_{\text{error}}$</td>
<td>1.74</td>
<td></td>
</tr>
<tr>
<td>$df_{\text{error}}$</td>
<td>495</td>
<td></td>
</tr>
</tbody>
</table>

$t = 3.37$

Table 5.9 Protected $t$ between 'long die 1+2' and 'shrt d+p 1+2'

<table>
<thead>
<tr>
<th></th>
<th>long die 1+2</th>
<th>shrt d+p 1+2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\bar{X}_1$</td>
<td>5.20</td>
<td>5.94</td>
</tr>
<tr>
<td>$n_1$</td>
<td>200</td>
<td>200</td>
</tr>
<tr>
<td>$\text{MS}_{\text{error}}$</td>
<td>1.52</td>
<td></td>
</tr>
<tr>
<td>$df_{\text{error}}$</td>
<td>297</td>
<td></td>
</tr>
</tbody>
</table>

$t = 6.00$

The above two tables show the values of $t$ are 3.37 for the comparison between 'shrt die 1+2' and 'auxlries 1+2' and 6.00 for that between 'long die 1+2' and 'shrt d+p 1+2'. Because of 495 and 297 degrees of freedom for our error term respectively, for both comparisons the $t$ is significant at $\alpha= 0.05$. We therefore conclude that subjects perform better on the monomorphemic sequences than on their bi-morphemic counterparts of the same length, as shown in Table 5.7. These comparisons seem to indicate that given equal
phonological length, a random collection of content words has a lighter memory load than a collection of the same number of content words accompanied by function words. We cannot tell whether this is just because of the extra morphemes or whether it is also in part because of the need to link the word pairs syntactically.

5.4 Discussion

This preliminary experiment does distinguish memory load of sequences with different syntactic structure, and also confirms that the memory load of random sequences with a fixed phonological length increases with the number of morphemes. Hence, function words, by virtue of being morphemes, do add to memory load, though their phonological length is relatively short.

Baddeley, Thomson & Buchanan (1975) found that lists consisting of long words, which take a relatively long time to pronounce, cannot be recalled as well as lists consisting of short words. Word-length effects can also be found in the results of this experiment; the average success rate for the two 'shrt dic' sequences composed of twelve words of 48 morae long, 58.37%, was better than that for the two 'long dic' sequences composed of twelve words of 60 morae long, 43.33%. (It should be however noted that 'shrt dic 1' had the poorest result (51.91% of successes) amongst the four sequences in question, which may be because 'shrt dic 1' was the first random sequence presented to the subjects so that they may have had difficulty to cope with the novel experience.) Whilst word-length effects emphasises phonological length, the results of this experiment showed that the morpho-syntactic structure of words also had effect on recall.

The results also show that other kinds of variation also affect memorability. For example, a sequence of random numbers, 'nmb+clsf', was easier to recall than any other random sequence used in the experiment. This is especially interesting because each of the twelve units to be recalled was morphologically more complex - each unit was a complex composed of the digit, some of which also contained the magnitude indicator, and the classifier - than a whole word. One possible reason for the better performance on this sequence is that subjects were allowed to use the Roman numbers, with which numerals can be written down much faster than with Hiragana.
letters. The use of the Roman numbers should have caused reductions in processing demands in working memory.

The result of 'scrn nat 2', the scrambled version of the natural sentence, 'nat sent 2' seems to demonstrate 'Miller effects' or chunking effects. Miller (1958) showed that grammatical items were better recalled than random items, hypothesising that subjects could keep in mind individual items when combined into larger chunks which would decrease the total number of psychological units. The performance of 'nat sent 2' (89.16% of successes) was superior to that of 'scrn nat 2' (58.83% of successes), although 'nat sent 2' was presented previously to 'scrn nat 2', so subjects should have benefited from Hebb effects (Hebb 1961) - memory for repeated items improve over short-term memory trials. The better result for 'nat sent 2' suggests that Miller effects, which should have affected the syntactically organised sequence, 'nat sent 2', should be a greater influence than Hebb effects on short-term memorability.

The serial order effects are as expected on the basis of the primacy and recency effects, though syntactic structure effects may be shown by distortions, e.g. the little dip at unit 9 in Graph 5.1. The relatively smooth curve in Graph 5.1 will provide an interesting base line for comparison with the more interesting curves reported in later experiments.

Finally, the significant difference amongst the three natural sentences may relate to the difference in their syntactic difficulty. As the following diagrams (5.2), (5.3) and (5.4) show, the relative syntactic difficulty in terms of distance increases in the order of 'nat sent 2', 'nat sent 3', 'nat sent 1'. This is seemingly reflected by subjects' performance, which declines in the same order. Hence, the difference in performance on the natural sentences might be due to the variations in dependency distance.
'Stepping into the premises of a stranger's house without permission, even though it is an unoccupied house, is an illegal act.'
'Sitting in the sofa and drinking the remaining morning coffee, I tried clearing my head a little.'

(5.4) 'nat sent 3': Average dependency distance = 0.916 (= 22/24)

'The girl next to me put her finger on my wrist and drew a shapeless and strange figure there.'
The next questions which need to be asked will be what the effects on memory load of syntactic structure itself are, and in particular whether dependency distance has an effect which can be separated from other influences such as word order, self-embedding and extraction.
6 Dependency distance on Japanese sentences

In the first pilot experiment reported in the previous chapter, the question was investigated whether differing types of strings of words could induce different loads on short-term memory. The results of the experiment showed memory cost for different random sequences of equal length; in general the memory cost for content plus function-word strings appeared to be greater than for those consisting of content words only. This seemed to favour analysis in terms of morphemes rather than moras. In this chapter we shall extend the study into syntax.

6.1 Experiments 2 and 3

Experiments 2 and 3 were undertaken to investigate the effect of dependency distance as a measure of working memory load and performance on immediate recall tasks. As explained earlier, dependency distance is the distance between words and their parents, measured in terms of intervening words, and we predict that longer dependency distance entails worse performance on immediate recall tasks. However it is important to stress that we do not claim that this is the only way in which syntactic structure influences performance.

Two experiments are reported below where eighty-four subjects in all were tested on immediate recall tasks designed to test this prediction above. These experiments should be regarded as a pilot study, and the results reported below will need to be confirmed by further experiments, such as the one reported in Chapter 7.

6.2 Materials and Procedure

Materials
The stimulus sentences consisted of a total of 8 sentences, whose phonological lengths were all fixed at 48 morae. All the sentences were also made up of familiar Japanese words and were chosen to produce differences in dependency distance.

Dependency distance is a simple measure of syntactic difficulty that counts the number of words between a word and its parent, in which each
intervening word is equally associated with a distance of 1. To illustrate how dependency distance is measured, the dependency analysis of the sentence (6.1) is given below.

(6.1)

The total dependency distance: 34 (= 14+10+6+4)
The average dependency distance: 1.7 (=34/20 words)

'My son Noboru sent me a postcard of a vineyard from Bordeaux in France at the beginning of last month.'

N.B. • Grammatical relations:  s = subject;  o = object;  c = complement;  a = adjunct
• Word classes: For nouns, verbs, adverbs, numeral nouns and classifiers, translated words are given.
  Particles are sbj (= subject), top (= topic), obj (= object) and ptc (= other types of particles).
  Postpositional particles are translated into English.
  Auxiliary verbs are past, causative, passive, desiderative, negative and hearsay.
• I shows a boundary between phrases which depend directly on the main or embedded verb, so that they can be arranged in any order within the clause without a significant change in meaning.
The dependency distance of (6.1) is measured by counting the number of words between a dependent and its parent. Thus, a dependent which is adjacent to its parent enjoys the minimum distance, zero. The sentential-root is also assumed to be associated with no distance.

As can be seen from (6.1) above, in order to calculate the total distance, there are four words to be considered which are separated from their parent, yokoshi ‘to send’:

- the two direction particles ni ‘to’ and kara ‘from’
- the place particle ni ‘at’
- the topic particle wa.

First, the direction particle ni is separated by four words from its parent yokoshi, which results in a distance of 4. Second, there are two more words watashi ‘I’ and ni ‘to’ in addition to the four words explained above separating the place particle kara from its parent, resulting in a distance of 6. Third, on top of these six separating words above, four more words intervene between the dependent ni ‘at’ and its parent verb yokoshi ‘to send’, which corresponds to its distance of 10. Finally, the topic marker wa is separated from its parent yokoshi ‘to send’ by fourteen words, ten words described above plus four additional words, so the particle wa is associated with a distance of 14.

The total dependency distance for a sentence is given by summing all the distances incurred in the sentence, so the total distance for (6.1) is calculated by adding up the four distances depicted above, resulting in the total distance of 34. The average dependency distance for a sentence is gained by dividing its total distance by the number of words in the sentence. Hence, the average distance for (6.1) is 1.7, corresponding to the total distance, 34 divided by the total number of words in (6.1), 20.

In the following tables 6.1 and 6.2, the characteristics and the dependency distance for all the sentences utilised in the experiments are summarised. As can be seen from the tables below, there are only two exemplar stimuli for each type of sentence, so the results will not be conclusive. The bare sentences and their dependency analyses will be provided in the relevant sections below.
Table 6.1 Characteristics for all the sentences in Experiments 2 and 3

<table>
<thead>
<tr>
<th>experiments</th>
<th>pairs</th>
<th>characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experiment 2: clause</td>
<td>1a+1b, 2a+2b</td>
<td>A noun is modified by a long relative clause.</td>
</tr>
<tr>
<td>Experiment 3: max-min</td>
<td>3a+3b, 4a+4b</td>
<td>2 nouns are modified by the same number of morphemes.</td>
</tr>
</tbody>
</table>

N.B. ‘a’ and ‘b’ prefixed by the same number are a minimal pair. ‘a’ has an unmarked word order, while ‘b’ is a reordered version of ‘a’.

Table 6.2 Dependency distance for all the sentences in Experiments 2 and 3

<table>
<thead>
<tr>
<th>experiments</th>
<th>no.</th>
<th>μ</th>
<th>w</th>
<th>σ</th>
<th>distance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ttl</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ave</td>
</tr>
<tr>
<td>Experiment 2: clause</td>
<td>1a</td>
<td>48</td>
<td>23</td>
<td>4</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>1b</td>
<td>48</td>
<td>23</td>
<td>4</td>
<td>55</td>
</tr>
<tr>
<td></td>
<td>2a</td>
<td>48</td>
<td>23</td>
<td>4</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>2b</td>
<td>48</td>
<td>23</td>
<td>4</td>
<td>52</td>
</tr>
<tr>
<td>Experiment 3: max-min</td>
<td>3a</td>
<td>48</td>
<td>20</td>
<td>2</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>3b</td>
<td>48</td>
<td>20</td>
<td>2</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>4a</td>
<td>48</td>
<td>20</td>
<td>3</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>4b</td>
<td>48</td>
<td>20</td>
<td>3</td>
<td>14</td>
</tr>
</tbody>
</table>

N.B. μ = number of morae
w = number of words
σ = number of syntactic chunks whose final words are dependents of either the main or the embedded verb
distance = dependency distance
ttl = total dependency distance
ave = average dependency distance

Procedure
Experiments 2 and 3 were administered in July 1998 up to January 1999. A total of 84 native Japanese individuals who had at least a high-school education participated in the experiments. Subject groups A, B and C involved 36, 38
and 10 subjects respectively. Of the 84 subjects, 74 who constituted Groups A and B were students at the Institute of International Education in London, and Group C made up of the rest of 10 subjects were my personal acquaintances; all the subjects had an average age of 23 years (range 18 - 37).

Each subject group was tested twice with an interval of 1 week; in the first week, the sentences chosen from one of the minimal pairs were used, and their corresponding sentences in the second week. All the subjects in each group who were tested in the first week turned up on the second week as well.

It is important to recognise the possibility of "long-term" Hebb effects (Hebb 1961) from Week 1 to Week 2 whereby performance would improve in Week 2 as a result of experience of the same words and lexical chunks in Week 1. To some extent these effects were controlled by counterbalancing the order of presentation, but the sentences in Experiment 3 were not counterbalanced so the results have to be interpreted with caution.

The sentences and their order of presentation introduced for each subject group are displayed in Table 6.3 below. Although Group A were given six sentences in total, four of which were different types, and 5 different types of sentences in all were tested on Groups B and C in each week, the ones which are not reported in this chapter due to experimental design flaws are omitted in the table.

<table>
<thead>
<tr>
<th>Table 6.3 Order of the presentation of the sentences</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Subject Group A</strong></td>
</tr>
<tr>
<td>week 1</td>
</tr>
<tr>
<td>2.</td>
</tr>
<tr>
<td>5.</td>
</tr>
<tr>
<td>6.</td>
</tr>
<tr>
<td>week 2</td>
</tr>
<tr>
<td>2.</td>
</tr>
<tr>
<td>5.</td>
</tr>
<tr>
<td>6.</td>
</tr>
</tbody>
</table>

N.B. The left-most column shows the number of the order of presentation.

163
Table 6.3 above shows which subjects participated in which experiments; Subjects belonging to Group A took part in the experiments laid out in the left column, i.e. Experiments 2 (clause) and 3 (max-min). Subject groups B and C participated Experiment 3 (max-min) in the middle and right columns respectively.

Subjects in a group of 9 to 12 were tested on the immediate recall task - the same one employed in Experiment 1 - where they were required to write down each sentence as soon as I had read it out, and were reminded to aim only for accuracy of recall. The average rate of presentation was 0.15 seconds per mora. For each sentence, approximately one mora of pause was placed after each syntactic chunk, which is marked by a vertical line in the syntactic analysis of each sentence. All the sequences were read only once in a flat fashion with standard Japanese pitch accent and enough time was provided after the presentation of each sentence so that all the subjects could complete their reproduction of the sentence. The subjects were allowed to use any Japanese scripts, hiragana, katakana or kanji, and also Arabic numerals for digits.

The stimulus sentences were scored on the basis of the words in the materials, whereby each word was scored either as a success or as a failure; for each success, 1 point was given, and for a failure, zero. A reproduction of a word was a success if it was either a verbatim copy, a synonym or abbreviation of the original, or a deviation which could be due to mishearing. Any other recall was scored as a failure. Displaced or reversed words were counted so that a sentence containing such items would gain as many successes as possible.

6.3 Overall results

The results for the two experiments, in which the total failures and successes for all the stimulus sentences were calculated, are presented in two ways: in terms of the percentages of failures and successes, and in terms of the number of failures for each word in the stimulus sentences. Tables 6.4, 6.5 and 6.6 show the percentages of correct words on each sentence at each time of testing in Groups A, B and C respectively.
### Table 6.4 Results for Subject Group A

<table>
<thead>
<tr>
<th>Subject Group A</th>
<th>week 1</th>
<th>no.</th>
<th>sort</th>
<th>failure</th>
<th>success</th>
<th>% correct</th>
<th>distance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2.</td>
<td>2b</td>
<td>clause</td>
<td>217</td>
<td>611</td>
<td>73.7</td>
<td>2.260</td>
</tr>
<tr>
<td></td>
<td>5.</td>
<td>1a</td>
<td>clause</td>
<td>59</td>
<td>769</td>
<td>92.8</td>
<td>0.695</td>
</tr>
<tr>
<td></td>
<td>6.</td>
<td>3b</td>
<td>max-min</td>
<td>149</td>
<td>571</td>
<td>79.3</td>
<td>0.450</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>week 2</th>
<th>no.</th>
<th>sort</th>
<th>failure</th>
<th>success</th>
<th>% correct</th>
<th>distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.</td>
<td>1b</td>
<td>clause</td>
<td>99</td>
<td>729</td>
<td>88.0</td>
<td>2.391</td>
</tr>
<tr>
<td>5.</td>
<td>2a</td>
<td>clause</td>
<td>120</td>
<td>708</td>
<td>85.5</td>
<td>0.695</td>
</tr>
<tr>
<td>6.</td>
<td>3a</td>
<td>max-min</td>
<td>51</td>
<td>669</td>
<td>92.9</td>
<td>0.450</td>
</tr>
</tbody>
</table>

### Table 6.5 Results for Subject Group B

<table>
<thead>
<tr>
<th>Subject Group B</th>
<th>week 1</th>
<th>no.</th>
<th>sort</th>
<th>failure</th>
<th>success</th>
<th>% correct</th>
<th>distance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5.</td>
<td>4a</td>
<td>max-min</td>
<td>62</td>
<td>698</td>
<td>91.9</td>
<td>0.700</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>week 2</th>
<th>no.</th>
<th>sort</th>
<th>failure</th>
<th>success</th>
<th>% correct</th>
<th>distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.</td>
<td>4b</td>
<td>max-min</td>
<td>126</td>
<td>634</td>
<td>83.4</td>
<td>0.700</td>
</tr>
</tbody>
</table>

### Table 6.6 Results for Subject Group C

<table>
<thead>
<tr>
<th>Subject Group C</th>
<th>week 1</th>
<th>no.</th>
<th>sort</th>
<th>failure</th>
<th>success</th>
<th>% correct</th>
<th>distance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5.</td>
<td>4b</td>
<td>max-min</td>
<td>51</td>
<td>149</td>
<td>74.5</td>
<td>0.700</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>week 2</th>
<th>no.</th>
<th>sort</th>
<th>failure</th>
<th>success</th>
<th>% correct</th>
<th>distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.</td>
<td>4a</td>
<td>max-min</td>
<td>24</td>
<td>176</td>
<td>88.0</td>
<td>0.700</td>
</tr>
</tbody>
</table>

N.B. The number on the left-most column shows the number of the order of presentation.

- failure = the number of failures
- success = the number of successes
- % correct = the percentage of successes
- distance = dependency distance

### 6.4 Experiment 2: clause

This experiment was chosen to examine the effects of variation in word order that also affects the dependency distance. Experiment 2 used sentences 165
where one of the dependents was modified by a long relative clause, resulting in a long dependency distance when it separates other dependents from the main verb. In the (a) sentences the resulting average dependency distance is much less than in the corresponding (b) sentences, so the main independent variable here is dependency distance.

**Materials**

(1a) yakyuu no booru o butsuke rare te migimimi no chouryoku o ushinat ta otouto tol boku wal densha del puuru nil it ta.

baseball of ball obj to hit passive ptc right ear of hearing ability obj to lose past younger with I top train by swimming to to go past brother

(1b) boku wal densha del puuru nil yakyuu no booru o butsuke rare te

I top train by swimming to baseball of ball obj to hit passive ptc pool

migimimi no chouryoku o ushinat ta otouto tol it ta.

right ear of hearing ability obj to lose past younger with to go past brother

‘I went to the swimming pool by train with my younger brother who has lost the hearing ability of his right ear by getting hit with a baseball ball.’

(2a) honyaku no genkou o uketoru tame ni kuronuri no rimujin de ki

translation of draft obj to receive sake for black of limousine by to come -varnished

ta josei tol kare wal jimusho del mizuwari o l non da.

past lady with he sbj office at whisky-and obj to drink past -water

(2b) mizuwari o l jimusho del kare wal honyaku no genkou o uketoru tame whisky-and obj office at he sbj translation of draft obj to receive sake -water

ni kuronuri no rimujin de ki ta josei tol non da.

for black of limousine by to come past lady with to drink past -varnished
In (1a) the noun, *otouto* 'younger brother' is modified by a long relative clause which is placed at the beginning of the sentence with three other chunks following it, and similarly for *josei* 'lady' in (2a). In contrast, the same nouns are positioned just before the sentence-final verb in the corresponding sentences (1b) and (2b), so the other dependents are separated by the long relative clause noun. This reordering results in a significantly higher dependency distance for (1b) and (2b) in comparison to that for (1a) and (2a).

The dependency analyses for the pairs of 'clause' (1) and (2) are shown below. The words that are underlined are those which subjects found most difficult to hold in memory; they will be referred to in the discussion of the results.
(1b)

boku  wal  densha  del  puuru  nil  yakyuu  no  booru  o  butsuke  rare  te
I  top  train  by  swimming  to  baseball  of  ball  obj  to  hit  passive  ptc

19  17  15

(2a)

migimimi  no  chouryoku  o  ushinat  ta  otouto  tol  it  ta.
right  ear  of  hearing  ability  obj  to  lose  past  younger  with  to  go  past
brother

honyaku  no  genkou  o  uketoru  tame  ni  kuronuri  no  rimuline  de  ki
translation  of  draft  obj  to  receive  sake  for  black  of  limousine  by  to  come
-varnished

168
Results and Discussion

Table 6.7 presenting the failures and successes for the pairs (1) and (2) shows that the sentences were given to the 36 subjects in Group A only, but the order of the presentation of the items was counterbalanced in this case with (1a) and (2b) in Week 1 and (1b) and (2a) in Week 2 (shown by the dot). Subjects' performances on both pairs were consistent with the prediction by dependency distance that (1b) and (2b) are syntactically more difficult to process than their counterpart, though it is hard to explain why (2) proved so much harder than (1).

Table 6.7  Results for the sentences (1) and (2)

<table>
<thead>
<tr>
<th>group</th>
<th>no.</th>
<th>failure</th>
<th>success</th>
<th>success %</th>
<th>distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1a</td>
<td>59</td>
<td>769</td>
<td>92.8</td>
<td>0.695</td>
</tr>
<tr>
<td></td>
<td>1b</td>
<td>99</td>
<td>729</td>
<td>88.0</td>
<td>2.391</td>
</tr>
<tr>
<td></td>
<td>2a</td>
<td>120</td>
<td>708</td>
<td>85.5</td>
<td>0.695</td>
</tr>
<tr>
<td></td>
<td>2b</td>
<td>217</td>
<td>611</td>
<td>73.7</td>
<td>2.260</td>
</tr>
</tbody>
</table>

169
Two-way Analysis of variance of a repeated-measures design shows that there was a difference between the two pairs, \( F(1,105) = 45.82, p < .05 \), and between the two sentence types (a) and (b), \( F(1,105) = 26.84, p < .05 \). There was also an interaction between pairs and types, \( F(1,105) = 4.65, p < .05 \), the reason for which will be examined below. Thus it seems that performance is better when the long relative clause is positioned so as to minimise dependency distance.

Table 6.8 displays the number of failures for each word in the sentences, in which the figure for individual words allow us to track the course of processing through these sentences. Table 6.8 below shows that subjects typically performed very differently on the (a) and (b) sentences.

There is a consistent relation across all the sentences between success and syntax, in that difficulty is greatest just after a word with a high dependency
distance. This can be seen more clearly in Graphs 6.1 and 6.2 and their accompanying dependency structures for the minimal pair (1).

Graph 6.1 Subject group A's performance on the sentence (1a)

As can be seen from Graph 6.1 above, in (1a) the number of failures for the first chunk (w1-15) modified by a long relative clause is consistently low except in the region w8-9, migimimi no 'of the right ear', which is after the high-distance word, w7, the particle te. After the end of the first chunk, w15, to 'with', whose distance is highest, 6, subjects' performance deteriorated on the following two short phrases boku wa 'I' in w16-17 and densha de 'by train' in w18-19.

As shown in Graph 6.2 below, the same pattern occurs in the reordered sentence (1b); subjects' performance declined in the region of w5-15, puuru ni yakyuu no booru o butsuke rare te migimimi no 'to the swimming pool, by getting hit with a baseball ball, of his right ear', which roughly overlaps with the position of the intervening long relative clause, w7-19, yakyuu no booru o butsuke rare te migimimi no chouryoku o ushinat ta 'who has lost the hearing ability of his right ear by getting hit with a baseball ball', and also is after the words bearing high dependency distance, w1-6, boku wa densha de puuru ni 'I, by train, to the swimming pool'.

171
Similarly comparable patterns appear in the other pair (2). Graph 6.3 above with its corresponding dependency structure below shows that in (2a) relatively worse performance can be observed after w7, ni ‘for’, and w15,
to 'with', which carry high dependency distance, although the worst performance occurs on w6, tame 'sake', just before the high-distance word, w7, for which I have no reasonable explanation.

Furthermore, Graph 6.4 below for the reordered sentence (2b) reveals a similar pattern to that for the reordered (1b); subjects' performance largely deteriorated in the region of w5-13, kare wa honyaku no genkou o uketoru tame ni 'he, to collect a draft of the translation', which roughly overlaps with the position of the intervening long relative clause, w7-19, honyaku no genkou o uketoru tame ni kuronuri no rimujin de ki ta 'who came in a black limousine to collect a draft of the translation' and also is after the words bearing high dependency distance, w1-4, mizuwari o jimusho de 'whisky-and-water, at the office'.

Graph 6.4  Subject group A's performance on the sentence (2b)
Comparing the four graphs above, we can see an overall similarity between each type; crudely speaking, there are two peaks in Graph 6.1 for (1a) and Graph 6.3 for (2a), and one big heap in Graph 6.2 for (1b) and Graph 6.4 for (2b). However, the profiles for the graphs differ in details, which presumably resulted in the significant interaction between pairs and types. For instance, the improvement of subjects' performance on the mid-region (w10-w15) in (2a) is not as stark as the one in (1a). Also the decline in performance on the long relative clause (w5-w13) for (2b) is more considerable and stretches longer than the one (w7-w13) for (1b). Hence, although there are differences in the degree of performance, the effects of dependency distance on memorability seems to be justified.

In summary, despite the small number of items tested in this experiment, its results showed that dependency distance in general stands as a valid measure of syntactic difficulty both in terms of the overall difficulty of the whole sentence and also in terms of the difficulty of the individual words.

6.5 Experiment 3: max-min

The sentences in this experiment were included in order to compare two different syntactic analyses explained below, the "maximal" and "minimal" analyses, so they are called 'max-min'. The paired sentences have the same dependency distance according to the maximal analysis but different distances measured in the light of the minimal analysis. However, after completing the experiment it became clear that this variable was confounded in the materials with the order of clause elements, so the results cannot be interpreted reliably and a separate experiment was needed (and is reported in Chapter 7).

Word Grammar and the lexicalist models (cf. Section 4.5) give rise to different morpho-syntactic analyses. The crucial difference between them lies in the treatment of "small words" such as particles and auxiliary verbs; every morpheme is treated as a separate word in the Word Grammar model, whereas the lexicalist analysis treats the particle or auxiliary verb as a mere suffix. These two approaches to Japanese syntax motivate two different syntactic analyses:
(i) the maximal analysis based on the Word Grammar approach, in which all the morphemes are considered as independent words, hence the maximum number of words are recognised in a given sentence.

(ii) the minimal approach, where a sentence is divided into the minimum number of words by identifying any function morphemes following a content word as parts of it.

One minimal-paired example below illustrates how the maximal and minimal analyses contribute to varying dependency distances for this type of sentence.

(3a)  

kurai monooki no dokoka de nakushi ta tokei o l

dark store-room of somewhere in to lose past watch obj

tetsudawa se rare taku nakat ta rashii imouto tol sagashimawat ta.

to help causative passive desiderative negative past hearsay younger with to search all over past sister

(3b)  

kurai monooki no dokoka de nakushi ta tokei o l sagashimawat ta.

dark store-room of somewhere in to lose past watch obj to search all over past

tetsudawa se rare taku nakat ta rashii imouto tol

to help causative passive desiderative negative past hearsay younger with sister

'I searched all over for the watch I had lost somewhere in the dark store-room with my younger sister who did not seem willing to be made to help.'

In these sentences, two dependent nouns of the sentence-root verb, namely tokei o 'watch' and imouto to 'with younger sister' in the pair (3) are modified by a relative clause. The order of the two dependent noun chunks
in (3a) is reversed in (3b). When morphemes are counted as separate words, the number of words as well as morae making up a relative clause is constant for the two paired sentences, i.e. 7 words and 16 morae, respectively. The syntactic structures of the sentences are comparable to each other; there are two relative clauses followed by a noun plus particle string and the sentential root verb at the end. The maximal and minimal dependency analyses for these sentences are given below.

**The maximal analysis applied to (3a)**

```
         a    c    a    c    a    c    a    o
kurai  monooki no dokoka de nakushi ta tokei ol
dark store-room of somewhere in to lose past watch obj
```

average dependency distance of (3a) = .450 (=9/20 words)

**The minimal analysis applied to (3a)**

```
         a    a    a    a    o
'c
kurai monooki-no dokoka-de nakushi-ta tokei-ol
dark store-room of somewhere-in to lose-past watch-obj
```

average dependency distance of (3a) = .250 (=2/8 words)
As can be seen from the syntactic analyses of the pair above, the maximal analysis predicts that the syntactic difficulty of both the sentences for each pair is the same, because their dependency distances are identical, 0.45. By contrast, the minimal analysis yields a different dependency distance for the pair; the distance of (3a), 0.250, is considerably shorter than that of its counterpart (3b), 0.625. This is mainly because of its treatment of the relative clause consisting of one full verb followed by six auxiliary verbs, which form one syntactic word: tetsudawa-se-rare-taku-nakat-ta-rashii 'did
not seem willing to be made to help'. Thus, the minimal analysis, in contrast to the maximal analysis, predicts that (3a) is easier to process than (3b). This type of sentence is relevant to exploring the appropriateness of the Word Grammar and lexicalist analyses as a measure of syntactic difficulty.

Materials
For the sentences (3a) and (3b) and their maximal and minimal analyses, see above.

(4a) kanamonoya de airon dai o kat ta boku wal byouin ni ika se
ironmonger’s at ironing board obj to buy past I top hospital to to go causative
rare ta otouto tol ikebukuroeki del ochiat ta.
passive past younger with Ikebukuro at to meet up past

(4b) byouin ni ika se rare ta otouto tol kanamonoya de airon dai o
hospital to to go causative passive past younger with ironmonger’s at ironing board obj
kat ta boku wal ikebukuroeki del ochiat ta.
to buy past I top Ikebukuro at to meet up past

'I who had bought an ironing board at the ironmonger’s met up at Ikebukuro station with my younger brother who had been made to go to hospital.'

The maximal and minimal analyses for the pair (4) are given below.
The maximal analysis applied to (4a)

The minimal analysis applied to (4a)
As mentioned earlier, the minimal pairs were constructed so as to compare the maximal and minimal analyses, but it later emerged that this contrast was confounded with another factor. Both (3a) and (4a) have the subject or
object before an adjunct, in contrast with the reversed order in (3b) and (4b). It will therefore be hard to know how to interpret the results. The results are nevertheless worth reporting because they are very clear, and a later experiment in Chapter 7 will help in the interpretation.

Results and Discussion

The number of failures and successes for the pairs (3) and (4) along with their dependency distances assessed by the maximal and minimal analyses are shown in Table 6.9. As before, the sentences presented in the second week are indicated by 

<table>
<thead>
<tr>
<th>Group</th>
<th>No.</th>
<th>Failure</th>
<th>Success</th>
<th>Success %</th>
<th>Maximal</th>
<th>Minimal</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>3a</td>
<td>51</td>
<td>669</td>
<td>92.9</td>
<td>0.450</td>
<td>0.250</td>
</tr>
<tr>
<td>A</td>
<td>3b</td>
<td>149</td>
<td>571</td>
<td>79.3</td>
<td>0.450</td>
<td>0.625</td>
</tr>
<tr>
<td>B</td>
<td>4a</td>
<td>62</td>
<td>698</td>
<td>91.9</td>
<td>0.700</td>
<td>0.666</td>
</tr>
<tr>
<td>B</td>
<td>4b</td>
<td>126</td>
<td>634</td>
<td>83.4</td>
<td>0.700</td>
<td>1.111</td>
</tr>
<tr>
<td>C</td>
<td>4a</td>
<td>24</td>
<td>176</td>
<td>88.0</td>
<td>0.700</td>
<td>0.666</td>
</tr>
<tr>
<td>C</td>
<td>4b</td>
<td>51</td>
<td>149</td>
<td>74.5</td>
<td>0.700</td>
<td>1.111</td>
</tr>
</tbody>
</table>

In this experiment, while the maximal analysis predicted that the sentences in each pair were equally difficult to process, the minimal analysis predicted that (3b) and (4b) were more difficult than their matching sentences. One-way analysis of variance of a repeated-measures design reveals that there were differences in performance between (3a) and (3b) in Group A as well as between (4a) and (4b) in Groups B and C (for the pair (3): $F(1,35) = 30.17, p < .05$; for the pair (4) $F(1,47) = 13.77, p < .05$). The results shown in Table 6.9 give tentative support to the prediction by the minimal analysis, because subjects' performance on (3b) and (4b) was consistently poorer than on their counterparts (3a) and (4a).

To summarise, this experiment showed clearly that (3a) and (4a) are easier than (3b) and (4b). Unfortunately these pairs differed in two respects: dependency distance as measured by the minimal analysis, and order of clause elements (S/O A for the (a) sentences, and A S/O for the (b) sentences). The results of Experiment 1 suggest that the minimal measure of dependency...
distance should not be relevant, so we may assume tentatively that the relevant influence is in fact the order of elements. However this question clearly needs more investigation, so an additional experiment was designed to test the minimal measure more specifically. This is reported in Chapter 7.

6.6 General discussion

The results of Experiment 2 (clause) are compatible with the prediction that dependency distance affects memory load. In this experiment, the increase in average dependency distance is caused by the intervening long relative clauses between the sentential verbs and their dependents. The results support the basic hypothesis about dependency distance: the longer the distance between a parent and its dependent is, the harder they are to process. By hypothesis, this effect is independent of any other effect that the particular constructions involved may have. However to confirm the existence of the relationship between dependency distance as a measure of working memory load and performance on immediate recall tasks, more extensive data collection is needed.

Performance data from Japanese in Hawkins (2001) and Yamashita (in press) reveal that there is an overall preference in production for placing a longer phrase before a short one rather than a short-before-long pattern. Dependency distance can account for this fact; the short-before-long pattern such as (1b) and (2b) in Experiment 2 causes higher distance than the long-before-short one as in (1a) and (1b), so presumably it should be more demanding on working memory to produce the short-before-long types of sentences. This experiment tentatively shows that the short-before-long pattern is more difficult to understand than the other pattern.

There seem to be some other interesting patterns which may be attributed to some syntactic effects on serial order. We can see from the four graphs for pairs (1) and (2) that on the whole, the serial position effects such as primacy and recency effects apply to the first and last "chunks" of the sentences. These effects however, extend over a surprisingly large number of words for if the words are linked with the dependency distance of zero (we will call this type of chunk a "chain"). The chains of more than two words in the initial and final positions are as follows:
As can be seen from Graph 6.1 to 6.4, the performance on these four chains is better than expected. (Certain words such as *migimimi no* in W14-W15 of (1b), *tame* in W6 of (2a) and *josei* in W20 of (2b) are exceptions; in fact, the performance deteriorates on these four words in both orders, so there may be possible item effects on them.) We shall term this lengthening of primacy and recency effects owing to chains "chaining" effects.

It should be noted that 'chain' is not equivalent to 'phrase' in Phrase-structure Grammar and also cannot be defined in phrase structure. Take the chains for example that contain the main-verb elements, W22-W23, *it ta* 'went' in pair (1) and *non da* 'drank' in pair 2. These chains, W20-W23 in (1a) or (2a) and W14-W23 in (1b) or (2b) form only a part of the sentence, whereas in phrase structure the phrase that holds the main-verb as its head would encompass the whole sentence.

Furthermore, we can observe other syntactic effects on the chains in the mid-region of the sentences in pairs (3) and (4). The chains in questions are as follows:

- W8-W15 in (1a).
- W7-W13 in (1b).
- W8-W15 in (2a).
- W7-W13 in (2b).

The chains in (1a) and (2a) have one unlinked dependency anchored at W7, whilst those in (1b) and (2b) are preceded by three unlinked dependency arrows anchored at W2, W4 and W6. If we compare the relevant four graphs, we can see that subjects' performance on the middle chains in (1a) and (2a) is relatively better than that on their counterparts in (1b) and (2b), respectively. (As mentioned earlier, W8-W9, *migimimi no*, in (1a) is the initial part of the middle chain, but the performance on them is relatively worse than the rest of the chain.) Thus, although a chain in general forms a syntactic chunk by virtue of zero dependency distance within it, it appears
that the performance on the chain deteriorates more if it is preceded by multiple unintegrated words. We will refer to this syntactic influence as "overload" effects, because this is the source of syntactic difficulty which could cause processing breakdown in such sentences with multiply nested structures as self-embedded ones.

The results of Experiment 3 (max-min) show some consistency of effects. The experiment appears to account for the way phonologically short words such as particles and auxiliary words are processed; according to the results which are in favour of the minimal analysis in terms of dependency distance, these function words are part of their preceding content words. However, as mentioned above, the results may be largely influenced by another factor, the order of clause elements (S/O A versus A S/O). Without more evidence it is impossible to choose between these interpretations. The word-order effect is tentatively shown by Experiment 4 in the next chapter.

In conclusion, Experiment 2 tentatively confirms a relationship between dependency distance as a measure of working memory load and subjects' performance on immediate recall tasks and such other syntactic effects as "chaining" effects, which lengthen primacy and recency effects on the chains in the initial and final positions, and "overload" effects, due to which the performance deteriorates more on the items subsequent to multiple unintegrated words. Experiment 3 apparently shows the effects of syntactic factors, such as the order of clause elements, which are independent of dependency distance and exert an influence on memory load.
In the previous chapter, several pilot tests were carried out to assess the relationship between a number of syntactic variables and performance on immediate recall tasks. The sentence type utilised in this experiment is equivalent to the one in Experiment 4 (max-min). The experiment was designed to investigate the appropriateness of measuring dependency distance in terms of the maximal and minimal analyses. As we shall see, this contrasts turned out to interact with word order differences in an interesting way.

7.1 Materials and Procedures

Materials. The stimuli consisted of 15 minimal-paired sentences, each of which contained 20 morphemes totalling 48 morae in length. As shown above, all the stimulus sentences had similar syntactic structures, i.e. the sentence-final verb has two dependents, each of which consists of the head, comprising a noun and particle, following a relative clause. To avoid an effect of the SOV/OSV word order, one of the two heads in a sentence was designed to be an adjunct of the main verb. The only difference between each pair was the order of the two relative clause plus head strings; one sentence in a pair had the reversed order of the other. All the relative clauses in the stimuli were constant in their number of morphemes and phonological length, namely 7 morphemes constituting 16 morae. The two relative clauses in a sentence differed in such a way that one clause contained 1 or 2 content words as opposed to the other accommodating 3 or 4. We shall call the ones in which a relative clause with more content words precedes the other Type A, e.g. (7.1a), and the others Type B, e.g. (7.1b). Table 7.1 below shows the dependency distance for all the stimuli utilised in the experiment, which are calculated according to both the maximal and minimal analyses. (All of the 30 sentences and their corresponding translations are listed in the Appendix at the end of this chapter.)
(7.1a)

kurai monooki no dokoka de nakushi ta tokei o
dark store-room of somewhere in to lose past watch obj

sagashimawat ta.

testsudawa se rare taku nakat ta rashii imouto to
causative passive negative past hearsay younger with

sagashimawat ta.
to search all over past sister

(7.1b)

testsudawa se rare taku nakat ta rashii imouto to
causative passive negative past hearsay younger with

sagashimawat ta.
sister

kurai monooki no dokoka de nakushi ta tokei o
dark store-room of somewhere in to lose past watch obj

sagashimawat ta.
to search all over past

'I searched all over for the watch I had lost somewhere in the dark
store-room with my younger sister who did not seem willing to be made
to help.'
Table 7.1  Dependency distance for all the stimuli in Experiment 4

<table>
<thead>
<tr>
<th>no.</th>
<th>max</th>
<th>min</th>
</tr>
</thead>
<tbody>
<tr>
<td>1a</td>
<td>.450</td>
<td>.333</td>
</tr>
<tr>
<td>1b</td>
<td>.450</td>
<td>.555</td>
</tr>
<tr>
<td>2a</td>
<td>.550</td>
<td>.555</td>
</tr>
<tr>
<td>2b</td>
<td>.550</td>
<td>.555</td>
</tr>
<tr>
<td>3a</td>
<td>.750</td>
<td>.666</td>
</tr>
<tr>
<td>3b</td>
<td>.750</td>
<td>.888</td>
</tr>
<tr>
<td>4a</td>
<td>.450</td>
<td>.285</td>
</tr>
<tr>
<td>4b</td>
<td>.450</td>
<td>.571</td>
</tr>
<tr>
<td>5a</td>
<td>.550</td>
<td>.375</td>
</tr>
<tr>
<td>5b</td>
<td>.550</td>
<td>.750</td>
</tr>
<tr>
<td>6a</td>
<td>.450</td>
<td>.250</td>
</tr>
<tr>
<td>6b</td>
<td>.450</td>
<td>.625</td>
</tr>
<tr>
<td>7a</td>
<td>.550</td>
<td>.500</td>
</tr>
<tr>
<td>7b</td>
<td>.550</td>
<td>.625</td>
</tr>
<tr>
<td>8a</td>
<td>.550</td>
<td>.444</td>
</tr>
<tr>
<td>8b</td>
<td>.550</td>
<td>.666</td>
</tr>
<tr>
<td>9a</td>
<td>.450</td>
<td>.250</td>
</tr>
<tr>
<td>9b</td>
<td>.450</td>
<td>.625</td>
</tr>
<tr>
<td>10a</td>
<td>.450</td>
<td>.333</td>
</tr>
<tr>
<td>10b</td>
<td>.450</td>
<td>.555</td>
</tr>
<tr>
<td>11a</td>
<td>.500</td>
<td>.285</td>
</tr>
<tr>
<td>11b</td>
<td>.500</td>
<td>.571</td>
</tr>
<tr>
<td>12a</td>
<td>.500</td>
<td>.400</td>
</tr>
<tr>
<td>12b</td>
<td>.500</td>
<td>.700</td>
</tr>
<tr>
<td>13a</td>
<td>.550</td>
<td>.375</td>
</tr>
<tr>
<td>13b</td>
<td>.550</td>
<td>.750</td>
</tr>
<tr>
<td>14a</td>
<td>.550</td>
<td>.444</td>
</tr>
<tr>
<td>14b</td>
<td>.550</td>
<td>.666</td>
</tr>
<tr>
<td>15a</td>
<td>.550</td>
<td>.500</td>
</tr>
<tr>
<td>15b</td>
<td>.550</td>
<td>.600</td>
</tr>
</tbody>
</table>

N.B.  no. = sentence number; Sentences numbered with the same sentence number are a minimal pair.
The suffixes 'a' and 'b' respectively refer to Type A and B.
max = dependency distance measured by the maximal analysis.
min = dependency distance measured by the minimal analysis.
The stimuli were composed of familiar words to represent easily comprehensible sentences. Each of the sentences was first recorded by a male speaker of standard Japanese in normal intonation with an average speech rate of approximately 6.8 morae per second. After every head string, one mora of pause was placed, which roughly corresponded to 0.15 seconds.

All 30 recorded sentences were first arranged into two sets: Set 1 and Set 2 as shown in Table 7.2 below. In each set, sentences were arranged according to the sentence numbers, from 1 to 15, and also alternated between Type A and B in order for these two types to be counterbalanced; Set 1 started with Type A and Set 2 with Type B.

<table>
<thead>
<tr>
<th>Set 1</th>
<th>Set 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1a</td>
<td>1b</td>
</tr>
<tr>
<td>2b</td>
<td>2a</td>
</tr>
<tr>
<td>3a</td>
<td>3b</td>
</tr>
<tr>
<td>4a</td>
<td>4b</td>
</tr>
<tr>
<td>5a</td>
<td>5b</td>
</tr>
<tr>
<td>6b</td>
<td>6a</td>
</tr>
<tr>
<td>7a</td>
<td>7b</td>
</tr>
<tr>
<td>8b</td>
<td>8a</td>
</tr>
<tr>
<td>9a</td>
<td>9b</td>
</tr>
<tr>
<td>10b</td>
<td>10a</td>
</tr>
<tr>
<td>11a</td>
<td>11b</td>
</tr>
<tr>
<td>12b</td>
<td>12a</td>
</tr>
<tr>
<td>13a</td>
<td>13b</td>
</tr>
<tr>
<td>14b</td>
<td>14a</td>
</tr>
<tr>
<td>15a</td>
<td>15b</td>
</tr>
</tbody>
</table>

Procedures. My subjects were eighteen undergraduate and graduate native Japanese students at University College London. All of my subjects heard all 30 sentences in one go; in order to counterbalance the two sets, a half of the subjects, we shall call them Group 1, were presented Set 1 first and then Set 2 second. Conversely, Set 2 was given before Set 1 to the other half, Group 2. In addition, prior to the sets subjects were presented with two dummy sentences, whose phonological length and number of words were identical to the stimuli.
The subjects took part in a series of small-scale experimental sessions which were conducted in the autumn and winter of 2000. The recorded sentences were presented over loudspeakers at a comfortable listening level at each session, where a group of between 1 and 3 students participated. Their task was to write each sentence down in any kind of Japanese script (hiragana, katakana or kanji) when it had been played. Subjects heard it only once and were given ample time for them to be able to write down whatever was held in their working memory. Because the purpose of this experiment was to test subjects' working memory retention, they were specifically instructed to aim for accuracy of recall.

The scoring system was based on morphemes: twenty morphemes in any given sentence were each scored either as a success or a failure. Subjects' reproduction of a morpheme was counted as a success and gained 1 point if:
• it was a verbatim copy of the original, or
• it was a synonym or abbreviation of the original, or
• discrepancy could be attributed to mishearing.
Any other response was marked as a failure, incurring zero points. Displaced or reversed morphemes were scored in such a way that they would receive as many successes as possible. For example, if the target is A B and the subject writes B A, then the B scores 0 but A scores 1.

7.2 Results

First, Table 7.3 and its accompanying graph give the total number of failures made by all the eighteen subjects for each morpheme in all the thirty sentences used in the experiment. Tables 7.4 and 7.5 show the total number of failures scored by each subject and scored for each sentence respectively.

Table 7.3 Number of failures for each morpheme in all the sentences

<table>
<thead>
<tr>
<th>serial number</th>
<th>w1</th>
<th>w2</th>
<th>w3</th>
<th>w4</th>
<th>w5</th>
<th>w6</th>
<th>w7</th>
<th>w8</th>
<th>w9</th>
<th>w10</th>
</tr>
</thead>
<tbody>
<tr>
<td>total</td>
<td>7</td>
<td>33</td>
<td>73</td>
<td>107</td>
<td>125</td>
<td>139</td>
<td>182</td>
<td>72</td>
<td>65</td>
<td>95</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>serial number</th>
<th>w11</th>
<th>w12</th>
<th>w13</th>
<th>w14</th>
<th>w15</th>
<th>w16</th>
<th>w17</th>
<th>w18</th>
<th>w19</th>
<th>w20</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>total</td>
<td>141</td>
<td>165</td>
<td>170</td>
<td>149</td>
<td>168</td>
<td>190</td>
<td>51</td>
<td>41</td>
<td>12</td>
<td>6</td>
<td>1991</td>
</tr>
</tbody>
</table>
Table 7.4  Total number of failures made by each subject

<table>
<thead>
<tr>
<th>Subject</th>
<th>Failures</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>215</td>
<td>35.83</td>
</tr>
<tr>
<td>S2</td>
<td>37</td>
<td>6.16</td>
</tr>
<tr>
<td>S3</td>
<td>286</td>
<td>47.66</td>
</tr>
<tr>
<td>S4</td>
<td>102</td>
<td>17.00</td>
</tr>
<tr>
<td>S5</td>
<td>119</td>
<td>19.83</td>
</tr>
<tr>
<td>S6</td>
<td>50</td>
<td>8.30</td>
</tr>
<tr>
<td>S7</td>
<td>109</td>
<td>18.16</td>
</tr>
<tr>
<td>S8</td>
<td>52</td>
<td>8.60</td>
</tr>
<tr>
<td>S9</td>
<td>70</td>
<td>11.66</td>
</tr>
<tr>
<td>S10</td>
<td>122</td>
<td>20.33</td>
</tr>
<tr>
<td>S11</td>
<td>101</td>
<td>16.83</td>
</tr>
<tr>
<td>S12</td>
<td>149</td>
<td>24.83</td>
</tr>
<tr>
<td>S13</td>
<td>24</td>
<td>4.00</td>
</tr>
<tr>
<td>S14</td>
<td>87</td>
<td>14.50</td>
</tr>
<tr>
<td>S15</td>
<td>115</td>
<td>19.16</td>
</tr>
<tr>
<td>S16</td>
<td>32</td>
<td>5.33</td>
</tr>
<tr>
<td>S17</td>
<td>165</td>
<td>27.50</td>
</tr>
<tr>
<td>S18</td>
<td>156</td>
<td>26.00</td>
</tr>
</tbody>
</table>

Total  1991  18.43
Average 110.61 18.43
<table>
<thead>
<tr>
<th>Sentence</th>
<th>Failures</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1a</td>
<td>99</td>
<td>27.50</td>
</tr>
<tr>
<td>1b</td>
<td>57</td>
<td>15.83</td>
</tr>
<tr>
<td>2a</td>
<td>57</td>
<td>15.83</td>
</tr>
<tr>
<td>2b</td>
<td>54</td>
<td>15.00</td>
</tr>
<tr>
<td>3a</td>
<td>61</td>
<td>16.94</td>
</tr>
<tr>
<td>3b</td>
<td>75</td>
<td>20.83</td>
</tr>
<tr>
<td>4a</td>
<td>55</td>
<td>15.27</td>
</tr>
<tr>
<td>4b</td>
<td>74</td>
<td>20.55</td>
</tr>
<tr>
<td>5a</td>
<td>57</td>
<td>15.83</td>
</tr>
<tr>
<td>5b</td>
<td>60</td>
<td>16.66</td>
</tr>
<tr>
<td>6a</td>
<td>86</td>
<td>23.88</td>
</tr>
<tr>
<td>6b</td>
<td>106</td>
<td>29.44</td>
</tr>
<tr>
<td>7a</td>
<td>52</td>
<td>14.44</td>
</tr>
<tr>
<td>7b</td>
<td>88</td>
<td>24.44</td>
</tr>
<tr>
<td>8a</td>
<td>76</td>
<td>21.11</td>
</tr>
<tr>
<td>8b</td>
<td>47</td>
<td>13.05</td>
</tr>
<tr>
<td>9a</td>
<td>78</td>
<td>21.66</td>
</tr>
<tr>
<td>9b</td>
<td>102</td>
<td>28.33</td>
</tr>
<tr>
<td>10a</td>
<td>107</td>
<td>29.72</td>
</tr>
<tr>
<td>10b</td>
<td>109</td>
<td>30.27</td>
</tr>
<tr>
<td>11a</td>
<td>68</td>
<td>18.88</td>
</tr>
<tr>
<td>11b</td>
<td>53</td>
<td>14.72</td>
</tr>
<tr>
<td>12a</td>
<td>31</td>
<td>8.61</td>
</tr>
<tr>
<td>12b</td>
<td>26</td>
<td>7.22</td>
</tr>
<tr>
<td>13a</td>
<td>33</td>
<td>9.16</td>
</tr>
<tr>
<td>13b</td>
<td>30</td>
<td>8.33</td>
</tr>
<tr>
<td>14a</td>
<td>76</td>
<td>21.11</td>
</tr>
<tr>
<td>14b</td>
<td>57</td>
<td>15.83</td>
</tr>
<tr>
<td>15a</td>
<td>53</td>
<td>14.72</td>
</tr>
<tr>
<td>15b</td>
<td>64</td>
<td>17.77</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>1991</td>
<td>18.43</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td>66.36</td>
<td>18.43</td>
</tr>
</tbody>
</table>
The overall average failure percentage for all subjects and all sentences was 18.43%, with the failure percentages per subject ranging from 4.00% to 47.66% and per sentence ranging from 7.22% to 30.27%.

Graph 7.1 mapping the total number of failures over the serial position of morphemes in the sentences confirms the serial position effects: the primacy effects on the initial region between w1 and w2 and the recency effects on the final morphemes w17 through w20. The noticeable dip in the mid region, w8 and w9 in the graph, will be discussed below.

Next, the tables below present the total numbers of failures for each set made by each group of subjects.

<table>
<thead>
<tr>
<th>Table 7.6</th>
<th>Total number of failures made by Group 1 (S1-S9)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Set 1</td>
<td>No. of failures</td>
</tr>
<tr>
<td>1a</td>
<td>61</td>
</tr>
<tr>
<td>2b</td>
<td>31</td>
</tr>
<tr>
<td>3a</td>
<td>47</td>
</tr>
<tr>
<td>4b</td>
<td>50</td>
</tr>
<tr>
<td>5a</td>
<td>44</td>
</tr>
<tr>
<td>6b</td>
<td>64</td>
</tr>
<tr>
<td>7a</td>
<td>39</td>
</tr>
<tr>
<td>8b</td>
<td>32</td>
</tr>
<tr>
<td>9a</td>
<td>43</td>
</tr>
<tr>
<td>10b</td>
<td>69</td>
</tr>
<tr>
<td>11a</td>
<td>53</td>
</tr>
<tr>
<td>12b</td>
<td>23</td>
</tr>
<tr>
<td>13a</td>
<td>30</td>
</tr>
<tr>
<td>14b</td>
<td>39</td>
</tr>
<tr>
<td>15a</td>
<td>37</td>
</tr>
<tr>
<td>Total</td>
<td>662</td>
</tr>
<tr>
<td>%</td>
<td>24.51</td>
</tr>
</tbody>
</table>
Table 7.7 Total number of failures made by Group 2 (S10-S18)

<table>
<thead>
<tr>
<th>Set 2</th>
<th>No. of failures</th>
<th>Set 1</th>
<th>No. of failures</th>
</tr>
</thead>
<tbody>
<tr>
<td>1b</td>
<td>24</td>
<td>1a</td>
<td>38</td>
</tr>
<tr>
<td>2a</td>
<td>38</td>
<td>2b</td>
<td>23</td>
</tr>
<tr>
<td>3b</td>
<td>44</td>
<td>3a</td>
<td>14</td>
</tr>
<tr>
<td>4a</td>
<td>35</td>
<td>4b</td>
<td>24</td>
</tr>
<tr>
<td>5b</td>
<td>40</td>
<td>5a</td>
<td>13</td>
</tr>
<tr>
<td>6a</td>
<td>53</td>
<td>6b</td>
<td>42</td>
</tr>
<tr>
<td>7b</td>
<td>63</td>
<td>7a</td>
<td>13</td>
</tr>
<tr>
<td>8a</td>
<td>59</td>
<td>8b</td>
<td>15</td>
</tr>
<tr>
<td>9b</td>
<td>62</td>
<td>9a</td>
<td>35</td>
</tr>
<tr>
<td>10a</td>
<td>68</td>
<td>10b</td>
<td>40</td>
</tr>
<tr>
<td>11b</td>
<td>31</td>
<td>11a</td>
<td>15</td>
</tr>
<tr>
<td>12a</td>
<td>14</td>
<td>12b</td>
<td>3</td>
</tr>
<tr>
<td>13b</td>
<td>16</td>
<td>13a</td>
<td>3</td>
</tr>
<tr>
<td>14a</td>
<td>55</td>
<td>14b</td>
<td>18</td>
</tr>
<tr>
<td>15b</td>
<td>37</td>
<td>15a</td>
<td>16</td>
</tr>
<tr>
<td>Total</td>
<td>639</td>
<td>Total</td>
<td>312</td>
</tr>
<tr>
<td>%</td>
<td>23.66</td>
<td>%</td>
<td>11.55</td>
</tr>
</tbody>
</table>

Tables 7.6 and 7.7 demonstrate that for both the groups, subjects’ performance improved on the later set of sentences that they were tested with; in terms of the percentages of the total failures, the figure for the former set is roughly halved for the latter. This is merely owing to the fact that the sentences of each minimal pair contained exactly the same words, which resulted in Hebb effects (Hebb 1961) on the experiment. However the experimental design, in which the orders of presentation was reversed for the two groups of subjects, allows us to ignore this effect in measuring the other effects.

Finally, the data is analysed using a repeated-measures two-way analysis of variance. The results showed that although there was a significant difference amongst subjects’ performance on all the fifteen minimal pairs, $F(14,493) = 7.85, p < .05$, there was no difference between Types A and B, $F(1,493) < 1$. This is as predicted by the maximal analysis, since their dependency distances were identical on this measure. There was however, an interaction between pair and type, $F(14,493) = 1.93, p < .05$, the reason for which will be explored in the following section.
7.3 Discussion

The major aim of the experiment was to evaluate the maximal and minimal analyses in order to decide which one best defines dependency distance as a measure of syntactic difficulty. On the one hand, the maximal analysis predicted that there would be no difference between Types A and B in terms of syntactic difficulty. But on the other hand, it was predicted by the minimal analysis that in general the syntactic difficulty of Type B should be greater than that of Type A (cf. Table 7.1 presenting the dependency distances for all the items.). The subjects' performance and the number of failures for each minimal pair show that the processing difficulty is relatively similar for the two types; the difference in the percentage of failures between the two types is within 6% in eleven out of fifteen pairs, namely pairs (2), (3), (4), (5), (6), (10), (11), (12), (13), (14) and (15) (cf. Table 7.5 displaying the percentage of failures for all the sentences.).

As to the four remaining pairs, pairs (1) and (8), and pairs (7) and (9), whose difference in the percentage of failures is more than 6%, subjects' performance on the two types move in opposite directions; for (1) and (8), the performance on Type A was better than on Type B, whereas Type B was better performed than Type A for (7) and (9). This indicates the interaction between pairs and types, mentioned above. Although there seems to be no conclusive reason for this, a tentative explanation is given in line with word-order.

Each sequence contained two syntactic heads which were direct dependents of the sentence-root verb, hence some of them served as such syntactically prominent functions as subject and object. The order of the heads as syntactic categories and the percentages of failures for all items are summarised below in Table 7.8.
<table>
<thead>
<tr>
<th>Sentence</th>
<th>Order of Heads</th>
<th>% Failures</th>
<th>% Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1a</td>
<td>A S'V</td>
<td>27.50</td>
<td>+11.67</td>
</tr>
<tr>
<td>1b</td>
<td>S'A V</td>
<td>15.83</td>
<td></td>
</tr>
<tr>
<td>8a</td>
<td>A S'V</td>
<td>21.11</td>
<td>+8.06</td>
</tr>
<tr>
<td>8b</td>
<td>S'A V</td>
<td>13.05</td>
<td></td>
</tr>
<tr>
<td>14a</td>
<td>S'A V</td>
<td>21.11</td>
<td>+5.28</td>
</tr>
<tr>
<td>14b</td>
<td>A S'V</td>
<td>15.83</td>
<td></td>
</tr>
<tr>
<td>11a</td>
<td>A S'V</td>
<td>18.88</td>
<td>+4.16</td>
</tr>
<tr>
<td>11b</td>
<td>S'A V</td>
<td>14.72</td>
<td></td>
</tr>
<tr>
<td>12a</td>
<td>A S'V</td>
<td>8.61</td>
<td>+1.39</td>
</tr>
<tr>
<td>12b</td>
<td>S'A V</td>
<td>7.22</td>
<td></td>
</tr>
<tr>
<td>2a</td>
<td>A S'V</td>
<td>15.83</td>
<td>+0.83</td>
</tr>
<tr>
<td>2b</td>
<td>S'A V</td>
<td>15.00</td>
<td></td>
</tr>
<tr>
<td>13a</td>
<td>A S'V</td>
<td>9.16</td>
<td>+0.83</td>
</tr>
<tr>
<td>13b</td>
<td>S' A V</td>
<td>8.33</td>
<td></td>
</tr>
<tr>
<td>10a</td>
<td>S'A V</td>
<td>29.72</td>
<td>-0.55</td>
</tr>
<tr>
<td>10b</td>
<td>A S'V</td>
<td>30.27</td>
<td></td>
</tr>
<tr>
<td>5a</td>
<td>A S'V</td>
<td>15.83</td>
<td>-0.83</td>
</tr>
<tr>
<td>5b</td>
<td>S' A V</td>
<td>16.66</td>
<td></td>
</tr>
<tr>
<td>15a</td>
<td>A S'V</td>
<td>14.72</td>
<td>-3.05</td>
</tr>
<tr>
<td>15b</td>
<td>S'A V</td>
<td>17.77</td>
<td></td>
</tr>
<tr>
<td>3a</td>
<td>A S'V</td>
<td>16.94</td>
<td>-3.89</td>
</tr>
<tr>
<td>3b</td>
<td>S'A V</td>
<td>20.83</td>
<td></td>
</tr>
<tr>
<td>4a</td>
<td>O A V</td>
<td>15.27</td>
<td>-5.28</td>
</tr>
<tr>
<td>4b</td>
<td>A O V</td>
<td>20.55</td>
<td></td>
</tr>
<tr>
<td>6a</td>
<td>S O'V</td>
<td>23.88</td>
<td>-5.56</td>
</tr>
<tr>
<td>6b</td>
<td>O'S V</td>
<td>29.44</td>
<td></td>
</tr>
<tr>
<td>9a</td>
<td>A A V</td>
<td>21.66</td>
<td>-6.67</td>
</tr>
<tr>
<td>9b</td>
<td>A A V</td>
<td>28.33</td>
<td></td>
</tr>
<tr>
<td>7a</td>
<td>S'A V</td>
<td>14.44</td>
<td>-10.00</td>
</tr>
<tr>
<td>7b</td>
<td>A S'V</td>
<td>24.44</td>
<td></td>
</tr>
</tbody>
</table>

N.B. % Failures = percentage of failures.
Order of Heads = order of heads.
S = subject; S' = topicalised subject; O = object, O' = topicalised object; A = adjunct; V = sentential verb.
% Difference = percentage of difference; the value is calculated by subtracting (b) from (a).
First of all, as can be seen from Table 7.8, the sequences were not counterbalanced in terms of the order of syntactic heads, which is because I had not considered this potential effect on syntactic processing. Concentrating on the data for the four pairs in question, i.e. (1), (7), (8) and (9), we can observe that in these pairs the sentence in which the topicalised subject preceded the adjunct was easier than the reverse order. This leaves (9), where both the dependents are adjuncts, to be explained. These three results may therefore be taken as tentative evidence that topics are easier to process when they are initial - an a priori plausible conclusion. The same pattern can also be seen in the pairs (2), (10), (11) and (12), although their differences are smaller than those of the previously mentioned three pairs.

Pairs (4) and (13), which have different types of word order, namely A S/O V versus S/O A V, show that the performance on the S/O A V ordered sentences was better than that on the A S/O V one. It is impossible to know whether there was any word-order effect on these sentences with so few samples.

It should be emphasised that the word-order effect concerning topicalisation is merely one of the other factors such as word-frequency/familiarity, word-length and chunking. Moreover the hypothesis about initial topics being easier is only weakly supported, because in pairs (3), (5), (14) and (15), the ASV ordered items were better performed than the SAV ordered ones, although the differences in performance on these pairs were rather small. I shall leave the effect of word-order on syntactic processing for future research.

The experiment tentatively demonstrated that dependency distance measured by the maximal analysis was more appropriate than that of the minimal analysis for the purpose of measuring syntactic difficulty. The results are relevant for theories of processing, as well as for theories of syntactic structure. At least for Japanese, memory load cannot be measured in terms of long morphologically complex "words"; rather the number of morphemes should be more relevant, therefore the Word Grammar syntactic analysis, where word and morpheme are equal, may be right.

Another question which arose as a by-product of this experiment was how to explain the effects of serial position within the string. Instead of the
even curve found in classic memory tests, the serial effects in Graph 7.1 show a very clear improvement in the middle of all the sentences. As explained previously, all the sentences shared a comparable syntactic structure where the first head was positioned at w8 and w9, the second one at w17 and w18, and the sentence-root verb at w19 and w20. These two heads were direct dependents of the final verb, therefore a syntactically well-formed sentence could be assembled from these three chunks. Take (7.1a') and (7.1b') for instance.

(7.1a')

As can be seen from the above, (7.1a') and (7.1b'), i.e. the extracted strings of (7.1a) and (7.1b) respectively, are simple OAV and AOV sentences with no relative clauses modifying the object or adjunct strings. Thus w8 and w9 in each string are the lexical and functional heads of the clause's first main phrase. This may explain why they were recalled so much more easily than the words before and after them, on the assumption that once a word is integrated semantically into a larger phrase, it can be discarded from working memory. In this rather unnatural test, a discarded word has to be
reconstructed from the phrase's meaning, which may lead to errors. But words w8 and 9 must be held in memory right up to the end of the sentence before they can be integrated, so they can simply be written down directly, without being reconstructed.

If this interpretation is right, the serial position curve (cf. Graph 7.1 displaying the total number of failures for each morpheme in the sentences) shows clearly that it is not just the number of morphemes affecting short-term memorability, but syntactic structure is also relevant. This is because the serial position curve is an indication that in Japanese it is the last word of the phrase that is preserved in working memory. Since the last word is also the phrase's head, we cannot distinguish two different interpretations of the data: one in terms of mere order (the last word in a phrase is the easiest to remember) and one in terms of dependency structure, as explained above, which we may call the "Head effect". The only way to decide between these analyses is to change to a language such as English where heads tend not to be phrase-final. If the Head effect holds, it can be predicted that the immediate recall for the translated English versions of the items utilised in this experiment will be better on the head segments. Take the English version of pair (4) for example.

(7.2)

The Head effect predicts that the following words, which are underlined in (7.2), will be better recalled: the subject, I, the sentence-root verb, searched,
and the two antecedent segments, *for the watch*, and *with my younger sister*. Whether this prediction is right will be left as a further research topic.

If the Head effect does turn out to be responsible for these processing differences, it will support the dependency analysis, where the head word stands semantically for the whole phrase, rather than the phrase structure analysis, where the head word has no special status. In the dependency analysis, the head word carries all the phrase's syntactic features and semantic information, so there is no need for a separate (mother) node for the whole phrase, whilst the phrase structure analysis recognises mother nodes in different bar-levels, so phrases are explicit in the phrase structure, but no distinction is given to their head words.
Appendix: Experimental materials for Experiment 4

The following are the experimental items used in the experiment. All thirty sentences are presented with their English translation.

(1a) karou de itame ta hai o chiryou-suru
overwork by to damage past lung object to treat
byouin ni arubaito-sa seru beki de wa
hospital at to do a part-time job causative conjectural with topic
nakat ta otouto wa nyuuin-shi ta.
negative past younger brother topic to be hospitalised past

(1b) arubaito-sa seru beki de wa nakat ta
causative part-time job causative with topic negative past
otouto wa karou de itame ta hai
younger brother topic overwork by to damage past lung
o chiryou-suru byouin ni nyuuin-shi ta.
object to treat hospital at to be hospitalised past

The younger brother who should not have been made to do a part-time job was admitted to the hospital that treats lungs damaged by overwork.'

(2a) michibata ni dosshiri to taore te iru
roadside on weary in to lie continual to be
tosainu ni omoiyari ga naku wa nakat
Tosa-dog to consideration subject negative topic negative

(2b) omoiyari ga naku wa nakat ta
consideration subject negative topic negative past

The younger brother who should not have been made to do a part-time job was admitted to the hospital that treats lungs damaged by overwork.'

200
'An old man who did not seem not to be considerate walked up to the Tosa-dog which was lying wearily on the roadside.'

(3a)

<table>
<thead>
<tr>
<th>tokidoki</th>
<th>tomodachi</th>
<th>de</th>
<th>yakyuu</th>
<th>o</th>
<th>yat</th>
<th>ta</th>
</tr>
</thead>
<tbody>
<tr>
<td>sometimes</td>
<td>friend</td>
<td>with</td>
<td>qbaseball</td>
<td>object</td>
<td>to do</td>
<td>past</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>yakyuujo</th>
<th>de</th>
<th>saibansho</th>
<th>ni</th>
<th>okuri</th>
<th>shime</th>
<th>rare</th>
</tr>
</thead>
<tbody>
<tr>
<td>baseball</td>
<td>feild</td>
<td>at</td>
<td>court of law</td>
<td>to</td>
<td>to send</td>
<td>causative</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ta</th>
<th>rou</th>
<th>shounen</th>
<th>wa</th>
<th>jisatsu-shi</th>
<th>ta</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>past</td>
<td>conjectural</td>
<td>boy</td>
<td>topic</td>
<td>to commit suicide</td>
<td>past</td>
<td></td>
</tr>
</tbody>
</table>

'A boy who would have been sent to the courts of justice committed suicide at a baseball ground where he sometimes played baseball with friends.'

(4a)

<table>
<thead>
<tr>
<th>kurai</th>
<th>monooki</th>
<th>no</th>
<th>dokoka</th>
<th>de</th>
<th>nakushi</th>
<th>ta</th>
<th>tokei</th>
</tr>
</thead>
<tbody>
<tr>
<td>dark</td>
<td>store-room</td>
<td>of</td>
<td>somewhere</td>
<td>in</td>
<td>to lose</td>
<td>past</td>
<td>watch</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>o</th>
<th>tetsudawa</th>
<th>se</th>
<th>rare</th>
<th>taku</th>
<th>nakat</th>
<th>ta</th>
<th>rashii</th>
</tr>
</thead>
<tbody>
<tr>
<td>object</td>
<td>to help</td>
<td>causative</td>
<td>passive</td>
<td>desiderative</td>
<td>negative</td>
<td>past</td>
<td>hearsay</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>imouto</th>
<th>to</th>
<th>sagashimawat</th>
<th>ta</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>younger sister</td>
<td>with</td>
<td>to search all over</td>
<td>past</td>
<td></td>
</tr>
</tbody>
</table>

(4b)

<table>
<thead>
<tr>
<th>tetsudawa</th>
<th>se</th>
<th>rare</th>
<th>taku</th>
<th>nakat</th>
<th>ta</th>
<th>rashii</th>
</tr>
</thead>
<tbody>
<tr>
<td>to help</td>
<td>causative</td>
<td>passive</td>
<td>desiderative</td>
<td>negative</td>
<td>past</td>
<td>hearsay</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>imouto</th>
<th>to</th>
<th>kurai</th>
<th>monooki</th>
<th>no</th>
<th>dokoka</th>
<th>de</th>
</tr>
</thead>
<tbody>
<tr>
<td>younger sister</td>
<td>with</td>
<td>dark</td>
<td>store-room</td>
<td>of</td>
<td>somewhere</td>
<td>in</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>nakushi</th>
<th>ta</th>
<th>tokei</th>
<th>o</th>
<th>sagashimawat</th>
<th>ta</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>to lose</td>
<td>past</td>
<td>watch</td>
<td>object</td>
<td>to search all over</td>
<td>past</td>
<td></td>
</tr>
</tbody>
</table>

'I searched all over for the watch I had lost somewhere in the dark store-room with my younger sister who did not seem willing to be made to help.'
(5a) kaisha no juuyaku ga atsuamru hiru no kaigi
company of executives subject to gather afternoon of meeting
de kaikeisya sa re nakere ba ike nakat
to solve passive negative conjunctive ought to do negative
ta jiken wa toriatsukawa reru.
past affair topic to deal with passive

(5b) kaikeisya sa re nakere ba ike nakat ta
kaikeisya sa to solve passive negative conjunctive ought to do negative past
jiken wa kaisha no juuyaku ga atsuamru
affair topic company of executives subject to gather
hiru no kaigi de toriatsukawa reru.
afternoon of meeting at to deal with passive

'An affair which should have been solved will be dealt with at an afternoon meeting where the company executives gather.'

(6a) kechi na yakunin no sakushu ni kurushimu
stingy copula officer of exploitation for to suffer
murabito wa omoidasa reru beki de
villagers topic to remember passive conjectural continuative
at ta youna dekigoto sae hanasa nai.
to be past conjectural incidents even to talk negative

(6b) omoidasa reru beki de at ta
to remember passive conjectural continuative to be past
youna dekigoto sae kechi na yakunin no
conjectural incidents even stingy copula officer of
sakushu ni kurushimu murabito wa hanasa nai.
exploitation for to suffer villagers topic to talk negative

'The villagers who suffer from the exploitation by the stingy officer never talk about even the incidents which should seem to have been remembered.'

(7a) yonaka kokudou o shissou-shi
the middle of the night national route object to dash
The driver who was dashing on the national route in the middle of the night died instantly due to carelessness which would not have arisen if he had not got drunk.'

'A Chinese person who could not be thought to want to go back worked at a harbour which was visible afar from the backyard of the house.'
(9a) 
ももたれがよくてきこんばしもよい

左手で振ってみるからすことはない。

(9b) 
よくて振ってみるからすことはない。

左手でももたれがよくてきこんばしもよい

(10a) 
耳にめんどうな不快を
手足の負担を

おやを支えておこしなさい。

(10b) 
おやを支えておこしなさい。

耳にめんどうな不快を

母に支えておこしなさい。

(11a) 
春の大学入試に

到かわる

母・父

高校

204
The professor who would not have been bothered attended a meeting concerning the university entrance examinations in the spring.'

'The child who did not seem to want to go home kept playing in the park where there were lots of yellow ginkgo trees.'
(13b)
renshuu-sa  se  nakere  ba  nara  nakat  ta
(to practise  causative  negative  conjunctive  must do  negative  past

seito  ga  senshuu  no  nichiyou  Yoyogi  de  at
pupil  subject  last week  of  Sunday  Yoyogi  in  to be

ta  konkuuru  de  nyuushou-shi  ta.
past  contest  at  to win a prize  past

'The pupil who should have been made to practise won a prize at the contest which was held in Yoyogi last Sunday.'

(14a)
Kyuushuu  kara  joukyou-shi  te  ma  mo  nai
(Kyuushuu  from  to move to Tokyo  continuative time  topic  negative

gakusei  wa  tashikame  sase  rare  taku  wa
student  topic  to make sure  causative  passive  desiderative topic

nai  youna  seiseki  ni  zetsubou-shi  ta.
negative  conjectural  mark  to  to despair  past

(14b)
tashikame  sase  rare  taku  wa  nai
to make sure  causative  passive  desiderative topic  negative

youna  seiseki  ni  Kyuushuu  kara  joukyou-shi
conjectural  mark  to  Kyuushuu  from  to move to Tokyo

te  ma  mo  nai  gakusei  wa  zetsubou-shi
continuative time  topic  negative  student  topic  to despair

ta.
past

'The student who had recently moved to Tokyo from Kyuushuu despaired at the marks which he did not want to be made to ascertain.'

(15a)
Shinjuku  no  ekimae  ni  aru  ooki  na
Shinjuku  of  front of station  in  to be  big  copula

depaato  de  shoutengai  ni  nakaro  u
department store  in  shopping centre  in  negative  future

206
The goods which should not be out of stock in the shopping centre cannot be found in a big department store which is situated in front of Shinjuku station.
8 Dependency distance in English and Japanese texts

Dependency distance was first developed and informally tested with immediate recall tests by Hudson (1995b, 1998), and he demonstrated its relevance to English. So far I have tentatively shown in Chapter 5 that in order to measure dependency distance each morpheme ought to be an independent unit for Japanese, and in Chapters 6 and 7 that dependency distance is also relevant to Japanese. Putting this into a broader perspective, we can address the next question: Given the same measure of syntactic difficulty applied to both languages, do Japanese speakers have to cope with greater difficulty than English speakers? This question arises because the assumption that long dependencies consume more mental energy seems to entail that Japanese is potentially more difficult than English, as will be explained in the following section.

This chapter introduces a small collection of performance data. By studying the products of performance - spoken English and Japanese texts in particular - we will examine how much strain is in fact placed on the working memory of speakers and hearers during the processes of ordinary speech production. By measuring difficulty in terms of dependency distance, it is possible to compare texts in the two languages.

8.1 Head-initial and -final language

Typological studies have revealed that languages in the world can be roughly classified as head-initial, head-final, mixed word order and word order-free on the basis of the order of a dependent (or a complement, specifier, or adjunct in Chomskyan phrase-structure grammar) in relation to a head (Crystal 1987: 98). Languages where heads consistently precede their dependents are referred to as head-initial languages (e.g. Welsh, Tongan, Squamish); whereas head-final languages such as Japanese and Amharic have dependents uniformly followed by their heads. Examples of mixed word order and free word order languages are Dutch and Navajo respectively. According to this classification, English falls into the mixed word order category, though linguists generally agree that English is predominantly a head-initial language, considering the fact that heads generally precede their dependents in the language in question.
The above four-way classification of language can be reduced to two: direction-consistent (e.g. VSO or SOV) and direction-inconsistent (e.g. SVO). Purely head-initial and head-final language are grouped as direction-consistent language, and mixed word order and word order-free language as direction-inconsistent. Thus, in dependency terms, Japanese is an example of the direction-consistent languages as opposed to English which would be included in the direction-inconsistent type.

One hypothesis concerned with cross-language processing is based on the plausible assumption that processing capacity is the same for all humans. This can be formulated as follows:

**The equal difficulty hypothesis**

There is no difference between direction-consistent and direction-inconsistent languages in their processing difficulties.

Taking Japanese and English as samples of direction-consistent and direction-inconsistent languages respectively, we can further hypothesise the following: Japanese is no more syntactically difficult than English. However, this seems to conflict with another hypothesis concerning the relationship between directionality and syntactic difficulty:

**The distance-difficulty hypothesis**

The sooner a word can be linked to its parent, the easier for the parser to process, i.e. the fewer words that separate them, the more undemanding for parsing.

The structural constraints on dependents located on one side of their parents would inevitably increase syntactic difficulty in comparison to bi-directional dependency structure. Take two simple three-word dependency structures below.

(8.1) 

(8.2) 

(8.3)
In (8.1), the bi-directional structure, the parent word \( W_3 \) is adjacent to its dependents \( W_1 \) and \( W_2 \) on both sides. The examples of mono-directional structure (8.2) and (8.3), on the contrary, have the parent \( W_3 \) only adjoining \( W_2 \) which in turn separates the other dependent \( W_1 \) from its parent \( W_3 \). Thus, theoretically speaking, (8.2) and (8.3) should be syntactically harder to process than (8.1) thanks to the presence of the intervening word \( W_2 \) between \( W_1 \) and \( W_3 \).

Since Japanese structures are overwhelmingly of type (8.2), Japanese potentially has more words than English that have to be held in working memory until they are linked with their parents, and therefore greater overall dependency distance. Consequently the hypothesis that Japanese is no more syntactically difficult than English presupposes that Japanese structures must have a smaller number of multiple dependents per word than comparable English ones do. Indeed, one characteristic in Japanese that would allow this assumption to be true is that any potential element in a Japanese sentence which would express information shared between the addresser and addressee is optional and therefore omissible. Consider the following example.

(8.4) \[ \text{boku [wa]} \ [\text{baaru itaria de}] \ [\text{koohii [o]}] \ [\text{kat ta}].^1 \]

I topic Bar Italia at coffee object buy past

'I bought some coffee at Bar Italia.'

First of all, for (8.4) to be grammatical, the sole requirement is that the root phrase of the sentence \( \text{kat ta} \) should be placed sentence-finally. Therefore, scrambling the other three items, each of which is formed with a noun followed by a particle, does not affect the well-formedness of the sentence. More importantly, in (8.4), any item in a bracket may be omitted if it is already understood between the speaker and the hearer. This omissibility makes Japanese quite dissimilar to English in syntactic structure, since in English obligatory subcategorisation frames specify the dependents of a verb. For example the English counterpart buy of the verb kau in (8.4) requires two dependents, specifically, a subject and an object. Therefore a

^1Omission of the place particle ‘de’ results in ungrammaticality as opposed to that of the topic particle ‘wa’ or the object particle ‘o’, although deletion of both ‘ga’ and ‘o’ leads to an ambiguous sentence in isolation. (‘Paul wani tabe ta.’ can be interpreted as ‘Paul ate a crocodile.’ or ‘A crocodile ate Paul.’).
sentence without them like ‘Bought at Bar Italia.’ is syntactically illicit, whilst the Japanese corresponding sentence ‘baaru itaria de katta.’ is still well-formed. In view of such syntactic structure, we can speculate that omissibility in Japanese is likely to favour the hypothesis that Japanese is no more syntactically difficult than English because it reduces the number of obligatory dependents in a Japanese sentence.

This speculation leads us to a third hypothesis which reconciles the first two: Speakers of a language with consistent word order tend to use fewer dependents per word than speakers of a language with inconsistent order. In the case of English and Japanese, Japanese texts must have a smaller number of multiple dependents per word thanks to omission than comparable English ones do. I shall explore in this chapter some English and Japanese spoken texts to see if they give evidence for this assumption.

8.2 Spoken and written language

 Whilst spoken language is acquired naturally, written language is taught at schools, therefore the contrast between them is relevant to the study of human societies and language (Goody 1997; Ong 1982). Furthermore the difference between spoken and written language is not only in the medium of communication, grammar or vocabulary but also in the general organisation of discourse or text (Miller 1994).

Based on various bodies of spoken language corpora such as Brown & Miller (1980) and Macaulay (1991), Miller (1994) claims that there is a significant number of syntactic constructions which occur in written English but are rare in spoken English. For instance, gapping, gerunds with subjects, and subordinate clauses introduced by although, since, and as are said to be uncommon in spoken English. He also reports that in spoken English information is spread over small “chunks” rather than integrated in one sentence.

In Japanese too the written and spoken language show different characteristics. As Shibatani (1990) states, one of the main differences between speech and the written language is formality; many linguistic features associated with informal-style Japanese disappear in the written language, although polite language is a feature of formal speech and informal
speech addressed to person of social status higher than the addressee. Another characteristic which is prevalent in spoken Japanese but absent in written Japanese is omissions of discourse particles such as *ne, yo, wa, sa* as well as grammatical particles such as the topic, subject, and object markers. Moreover - and most relevantly for us - the level of ellipsis is higher in spoken Japanese than in written Japanese.

Spoken and written language has different systems for different tasks. One advantage of studying spoken texts is that it can offer a faithful record of performance during language communication, since it is produced in real time, impromptu, with no time to edit. Also, spoken language is more relevant to short-term memory, because spontaneous speech is subject to the limitations of short-term memory. In contrast, written texts are advantageous for studying competence of language, or sentence grammar, because writing allows pauses for thinking and rewriting so as to produce more integrated, manipulable, complex syntax.

On the other hand, loosely structured, spoken texts can be problematic to analyse in terms of sentences, because they are hard to recognise in spoken language, especially conversation with interruption and overlapping turns. However phrases and clauses can typically be recognised unequivocally, so it is appropriate to treat spoken texts as collections of phrases and clauses.

My main concern in this chapter, as mentioned above, is whether Japanese speakers have to cope with greater difficulty than English speakers (i.e. pressure on working memory during production and understanding of language). Therefore it is important to compare similar kinds of texts, and it would be quite wrong to compare, say, written texts in one language with spoken texts in the other. I here present an analysis of five spoken texts as a pilot study. The data drawn from the five texts need to be treated with caution since this is rather a small sample in view of the fact that speech content varies depending on genres or registers, and it was impossible at the data gathering stage to control topic and setting, let alone speakers.

### 8.3 Materials and Method

*Materials.* In total five short texts were analysed: Two English (i.e. ET1 and ET2), and three Japanese (i.e. JT1, JT2 and JT3). All the texts can be
found in the Appendix at the end of this chapter. All texts are transcriptions of extracts about 500 words long from free conversations. ET1, ET2, JT1, and JT3 were produced and transcribed by students at the Phonetics and Linguistic department of University College London originally for the purpose of sociolinguistic research, and JT2 was arranged and transcribed by myself. The participants in the conversations were all native speakers of the language concerned. Also all the subjects were university students, with the exception of JT3 for which a Japanese lecturer took part in the conversation. Table 8.1 below displays the main characteristics of the texts. The data collection procedures were as follows: A free conversation was initially recorded onto a cassette tape, and some section from the entire discourse was transcribed verbatim and as accurately as possible.

<table>
<thead>
<tr>
<th>Text</th>
<th>Subjects</th>
</tr>
</thead>
<tbody>
<tr>
<td>ET1</td>
<td>4 English University students</td>
</tr>
<tr>
<td>ET2</td>
<td>4 English University students</td>
</tr>
<tr>
<td>JT1</td>
<td>6 Japanese University students</td>
</tr>
<tr>
<td>JT2</td>
<td>3 Japanese University students</td>
</tr>
<tr>
<td>JT3</td>
<td>3 Japanese University students plus a lecturer</td>
</tr>
</tbody>
</table>

**Method.** The analysis was done by calculating the mean dependency distance for each text.\(^2\) Firstly, speech error in the texts such as repeated words, hesitation words or sounds, and ungrammatical phrases were bracketed in order not to be analysed subsequently. Secondly, the dialogues in the texts were separated into independent syntactic groups. I shall simply use the term ‘phrase’ to refer to a separated unit. Thirdly, surface dependency analysis was applied to each phrase and the dependency distance for every word in the phrase was measured. Finally, the mean distance for a text was calculated by dividing its total distance, the sum of all the distances in the

\(^2\)Simple baseline analyses such as sentence / utterance length, type-token ratio are not carried out, since the analysis does not focus on difficulty that arises out of the physical length of sentence or the range of vocabulary but rather aims to compare the texts solely on the basis of difficulty resulting from syntax. In any case sentences and even utterances are very hard to identify in spoken texts.
text, by the total number of words in the text.

With regard to the segmentation and classification of words, the English analyses were primarily based on the Word Grammar Encyclopedia (Hudson, 1997b). As to the Japanese texts, the utterances were separated into words and classified according to section 4.4, and syntactically analysed by the maximal analysis justified in Chapter 7.

In determining dependency distance, any single word in the texts was counted equally, as one. The Japanese texts were analysed using the maximal analysis discussed in the previous chapter. The syntactically analysed texts were analysed quantitatively to produce a number of measures. In addition to the distance for each word, the proportion of parent words that had more than one dependent was also calculated.

8.4 Results

Table 8.2 represents the number of phrases, number of words, and gross and mean dependency distances of each text.

<table>
<thead>
<tr>
<th>Text</th>
<th>No. of Phrases</th>
<th>No. of Words</th>
<th>Gross Distance</th>
<th>Mean Distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>ET1</td>
<td>94</td>
<td>496</td>
<td>176</td>
<td>0.354</td>
</tr>
<tr>
<td>ET2</td>
<td>109</td>
<td>499</td>
<td>218</td>
<td>0.436</td>
</tr>
<tr>
<td>Total</td>
<td>203</td>
<td>995</td>
<td>385</td>
<td>0.386</td>
</tr>
<tr>
<td>JT1</td>
<td>106</td>
<td>501</td>
<td>187</td>
<td>0.373</td>
</tr>
<tr>
<td>JT2</td>
<td>94</td>
<td>498</td>
<td>177</td>
<td>0.355</td>
</tr>
<tr>
<td>JT3</td>
<td>105</td>
<td>496</td>
<td>244</td>
<td>0.491</td>
</tr>
<tr>
<td>Total</td>
<td>305</td>
<td>1495</td>
<td>609</td>
<td>0.407</td>
</tr>
</tbody>
</table>

A one-way analysis of variance with the dependency distance score for each word in the texts as observations shows that the difference between the two languages is not significant, $F(1,2488) < 1$. Thus so far as these five texts are concerned, we have no reason to suspect a difference in syntactic difficulty between English and Japanese.
Table 8.3 below demonstrates (i) the number of words possessing more than one dependent, (ii) the total number of words, (iii) the ratio between (i) and (ii).

Table 8.3 Results for words with multiple dependents

<table>
<thead>
<tr>
<th>Text</th>
<th>No. of words with &gt; 1 dep.</th>
<th>Total no. of words</th>
<th>(i)/(ii) ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>ET1</td>
<td>105</td>
<td>496</td>
<td>0.211</td>
</tr>
<tr>
<td>ET2</td>
<td>110</td>
<td>499</td>
<td>0.220</td>
</tr>
<tr>
<td>total</td>
<td>215</td>
<td>995</td>
<td>0.216</td>
</tr>
<tr>
<td>JT1</td>
<td>52</td>
<td>501</td>
<td>0.103</td>
</tr>
<tr>
<td>JT2</td>
<td>68</td>
<td>498</td>
<td>0.136</td>
</tr>
<tr>
<td>JT3</td>
<td>80</td>
<td>496</td>
<td>0.161</td>
</tr>
<tr>
<td>Total</td>
<td>200</td>
<td>1495</td>
<td>0.133</td>
</tr>
</tbody>
</table>

The ratio (iii) on Table 8.3 reveals that the Japanese texts have fewer words with multiple dependents; approximately 87 per cent of the words in the entire texts were linked in a one-to-one fashion with their parents, which was about 10 per cent more than for the English texts. The chi-square test applied to the contingency table 8.4 below showing the total number of words with one dependent and that with more than one dependent for all the English and Japanese texts confirms that the difference of the number of the words with multiple dependents between the English and Japanese texts is significant, $\chi^2(1, N= 2490) = 29.135, p < .05$.

Table 8.4 Data for words with one or multiple dependents

<table>
<thead>
<tr>
<th>Text</th>
<th>ET1+2</th>
<th>JT1+2+3</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of words with 1 dep.</td>
<td>780</td>
<td>1295</td>
<td>2075</td>
</tr>
<tr>
<td></td>
<td>(829.17)</td>
<td>(1245.83)</td>
<td></td>
</tr>
<tr>
<td>No. of words with &gt; 1 dep.</td>
<td>215</td>
<td>200</td>
<td>415</td>
</tr>
<tr>
<td></td>
<td>(165.83)</td>
<td>(249.17)</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>995</td>
<td>1495</td>
<td>2490</td>
</tr>
</tbody>
</table>

N.B. Expected frequencies in parentheses.
8.5 Discussion

The results show that the English and Japanese texts were statistically similar in terms of mean dependency distance, which indicates that English and Japanese speakers cope with a similar degree of difficulty in speech.

Text JT3 is interesting because it shows the highest mean distance amongst the five texts. This result seems to show that stylistic difference is relevant to dependency distance, because this text, a conversation between a lecturer and students, is more formal in style and closer to written language than any other Japanese texts. The other Japanese texts in contrast are remarkably similar in dependency distance to the socially similar English texts.

The ratio between the number of words with more than one dependent and the total number of words for each text confirms that Japanese spoken texts contain simpler syntactic structures with fewer words enjoying multiple dependents. Interestingly, this again seems to be influenced by formality, as JT3 is significantly higher than the other Japanese texts, \( \chi^2(1, N=1495) = 4.848, p < .05 \) (cf. Table 8.5). This means that the greater dependency distance in JT3 must also be due to more multiple dependents overall rather than a few very long dependencies.

<table>
<thead>
<tr>
<th>Text</th>
<th>JT1+2</th>
<th>JT3</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of words with 1 dep.</td>
<td>879</td>
<td>416</td>
<td>1295</td>
</tr>
<tr>
<td></td>
<td>(865.35)</td>
<td>(429.65)</td>
<td></td>
</tr>
<tr>
<td>No. of words with &gt; 1 dep.</td>
<td>120</td>
<td>80</td>
<td>200</td>
</tr>
<tr>
<td></td>
<td>(133.65)</td>
<td>(66.35)</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>999</td>
<td>496</td>
<td>1495</td>
</tr>
</tbody>
</table>

N.B. Expected frequencies in parentheses.

A sample of three phrases is demonstrated below: (8.5) and (8.6) from the texts of English and Japanese, respectively, and (8.7) selected from JT3. We can observe especially from (8.6) that singly-bounded dependencies occur to a great extent in the Japanese example in comparison to the English
it wouldn't have made [a great] any difference to me apart from if you boil-wash wool (taken from ET1) mean distance = 0.500 (7/14)

isshou no uchi des yon juu hachi kiro o koe ta koto ga nai lifetime of within in four ten eight kilo obj exceed past that sbj not exist (taken from JT1) mean distance = 0.312 (5/16)

'Your weight has never exceeded 48 kilos in your whole life, has it?'

gakubu wa dakara watashi wa ima wa oshie te mase n kara department top so I top now top to teach ptc pol neg you see (taken from JT3) mean distance = 0.916 (11/12)

'So, now I am not teaching in the department, you see.'
The data provided in this chapter support two very tentative conclusions: The one is that spoken Japanese is no harder to process than spoken English, if difficulty is measured in terms of dependency distance; and the other is that Japanese speakers keep dependency distance down to the English levels by using fewer words with multiple dependents. This conclusion will need a great deal more confirmation by empirical studies of texts. If it turns out to be true, it will confirm the close link between processing demands (minimising dependency distance) and grammatical structure (optimality) which has been the theme of functionalist work such as Hawkins (1994, 2001).
Appendix: Five transcriptions used for the text analysis

Transcription ET1 by Luisa Element 1995

T we had to do a project on washing
different fabrics, at different temperatures
1995

T we'd do a lot;/ / /*

no/<but/> it wouldn't have made (a great any difference to me
apart if you boil-wash wool/<>/, /

(and) that was it/<>

C what/it didn't make any difference at all
what about if you wash your whites
what about if you

T yeah/that shrinks/<>

C well/what if you boil-wash whites/<>/

T then, they go grey/<>/

C they don't (go white) get whiter/<>

T no/<if you>/ if you wash (erm) underwear at over forty degrees it eventually goes grey

T that's why white underwear goes grey/2/26

<T>/ <if you always hand-wash your underwear it'll never go grey/<>/

219
<C> yeah/<but> if you use colours, 4/4 = 0.000

<T> I used to never <like> put underwired bras in the washing machine, 5/5 = 0.00.

<D> <what's> what's the best way of getting all the white lines out of your collars and your cuffs? 9/9 = 0.00.

<T> well/I dunno/get the Vanish on them, 9/9, 2/5 = 0.00.

<J> Radion / 0/0.

<C> yeah/I think Vanish, 0/0, 0/3.

<J> what's wrong with <erm> putting underwired bras in, 4/8 = 0.00.

<C> yeah/what's wrong with that, 0/0, 1/5 = 0.00.

<T> cause it leaves a mark, it makes the bits pop out, 3/3 = 0.00.

<C> yeah/it's sort of, 0/0, 0/4.

<J> <and> you can't get the wire in, 2/6 = 0.333.

<J> <and> then the other gets you and stabs you in the chest, 6/14 = 0.66.

<T> <I mean> the most of the time it won't, 5/5 = 0.714.

<C> I've only ever had it happen, 6/9 = 0.666.

<T> well/you should always hand-wash your bra, 6/9, 6/6 = 0.16.

<C> I know that cause I did bra-training at [?], 3/9 = 0.333.

<J> I had to train everyone in how to fit a bra, 0/0, 0.09.

<T> <and> I had to train all the boys as well, 4/9 = 0.44.

220
T: (and) they were so embarrassed, it was like you know... fourteen year, fifteen year old boys doing bras about bras.

D: what? (and) you've got training to be able to fit bras.

T: yeah/not.

D: not for [?]

<but because if like>, if there aren't girls around then they can tell the lady how to go away and do it herself, hire a tape measure, they can say.

<but like>, they go so red.

<I mean>, can you imagine saying, measure around the fullest part of your bust.

(and) they're like [shshsh ha ha].

(and) that's the cup size.

C: <you> you're not on that any more are you.

T: no/I'm not, no-one's ever asked me to measure a bra in three and a half years...
T: I don’t know why I’ve never been trained that. I mean, you don’t even see the people in Littlewoods doing it or Marks-and-
Spencer. I mean, they’re supposed to be really into it, but it was really sad. I had to measure the department manager, yeah.
And she’s like, out there. She’s got fangs. Everyone was getting it wrong, you know. Saying things like, “thirty-eight double A”.
Like, “thirty-eight double A” would be sort of out there and flat. Sort of thing. I mean, you don’t even see the people in Littlewoods doing it or Marks-and-
Spencer. I mean, they’re supposed to be really into it, but it was really sad. I had to measure the department manager, yeah.
And she’s like, out there. She’s got fangs. Everyone was getting it wrong, you know. Saying things like, “thirty-eight double A”.
Like, “thirty-eight double A” would be sort of out there and flat. Sort of thing.
it would be wider and flatter

I'm gonna start that one

(hopefully which is you may) I don't

so I finish college

my final deadline's the sixteenth of March

(and) that's it for this year

(and) then (erm ha) that starts in May

so I've got nothing on afterward
I haven't got that really stale taste in your mouth.

No, I've just got like hungy.

You what?

Whenever I get up really early, I always feel angry.

Yeah.

I got up at half eight again this morning.

Why?

Ready for half past six. I just couldn't.

The last four mornings I got up at half eight.

I had to get out of bed.

I was awake.

Which is really not like me at all.

What's not like you?

To get up early.

To wake up at half eight.

And just get dressed.

What did you do that for?

I've done it for the last three or four mornings.
E 〈'n'〉 she comes down 'n' tidies up</n' that〉/α/β = 0.333
can't you see*/γ/δ = 0.333

Jy oh/yeah/α/β, α/β

A I did it yesterday morning/γ/δ = 0.25
I came down 'n' tidied up/γ/δ = 0.333
swear to god (man)/γ/δ

E when're your mum and dad coming, Anna/α/β = 0.75, α/β

A tomorrow/α/β

E you spoken to them/α/β

A tomorrow/α/β

Jo Anna/stop cheating/α/β it was my go/α/β, α/β, α/β

A d'you want me to shuffle them for you/γ/δ = 0.333

Jy oh/I can't wait to go home for reading week/α/β, γ/δ = 0.5

E are you going home all week/β/δ = 0.333

Jy yeah/well I don't know/α/β, α/β, 0/3
I've gotta wait 'n' speak to my dad tonight, cos I'm not sure/α/β

(0/7)/γ/δ = 0.937

A I'm going home tomorrow/γ/δ = 0.25

E oh/you've got James/α/β, α/β

Jy he doesn't know if he can pick me up/α/β = 0.25

225
E I'll speak to my dad cos he's coming down to London soon 9/3/22

Jy is he 9/3

Jo oh my God my dad's in London today 9/3/23 9/6 0:50

E+A is he 9/3

A have you seen him 9/4 0:50

Jo no I tried (but [?]) and then he's got to get back

because it is his mum and dad's best friend's twenty-fifth wedding anniversary 9/9 0:1 8/2 1:4 2:7

A wow 9/4

Jy it's my mummy's 5/9 0:6 0

Jo (and) so they're taking them out for a meal tonight 5/9/10 0:9 0

A ahh isn't that lovely 9/1 8/3 0:3 3

Jo so he's got to rush back from London 8/9 0:2 8

Jy ([?]) such a lovely weekend 9/4 0:2 5

A that's silver init 9/4 0:2 5

Jy yeah that's [?] 9/1 8/3

E my mum and dad's 4/9 0:8 0

when's your mum and dad's wedding anniversary 9/9 0:1 0 0

A is it 9/1

on my mum and dad we been (twenty-eight years) twenty-nine years 9/1 8/1 5/8 0:6
E 〈?〉 my mum and dad's is on the seventeenth of February/5/4/ = 0.654

(and) they're going away/0/4

A imagine sticking with someone for twenty-five years/9/4/ = 0.428

E that's this weekend innit/3/ = 0.400

when's the seventeenth/0/4

thirteenth, fourteenth, fifteenth, sixteenth/0/4, 0/4, 0/4, 0/4

that's Friday isn't/4/ = 0.850

they're going away this weekend/4/ = 0.66

Jo again/0/4

E yeah/0/4

Jo wahey/0/4

E just going to Stanford/0/4

is it George at Stanford/4/ = 0.466

Jy yep it's gorgeous< sort of> house /0/4, 3/5/ = 0.600

E it's just like this, really nice old hotel/5/8/ = 0.685

Jy it's really nice /4/ = 0.250

Jo 〈?〉 mine send them somewhere/4/ = 0.250

E yeah/get them out of the way/0/4, 4/ = 0.166

Jo 〈but〉 it's just where/4/ = 0.250

A d'you /0/4

Jo yeah/0/4

227
E what about /0/2

A d'you know what we should do/ 2/8 = 0.285

Jy what /0/A

A we should get our sisters down this week/ 5/8 = 0.645

Jy what/ I haven't got one/ 9/8, 0/4

I'm sorry/ 0/3

A no/me and Emily /9/8, 2/8 = 0.335

we should get our sisters down/ 0/6 = 0.333

Jo why /0/A

E I've gotta ring my sister tonight anyway / 5/9 = 0.555

A cos I haven't seen her for ages/ 2/7 = 0.285
Transcription JT1 by Izumi Nakamura 1994

Y kare wa san kiro yase ta n dak ke to 2/3

Mu yase ta no 0/3

H un yase ta yase ta 0/1, 0/2, 0/6

Y zutto ryou de gohan tabe te te 4/7 = 0.571

Ma igirisu-meshi da kara da yoon 0/5

Y kanojo wa hutot ta n dat te hachi kiro mo 0/7, 0/3

I ett hachi kiro 0/1, 0/6

Ma uso dou yat te hachi kiro mo hutoru no 0/1, 3/8 = 0.375

I sonna mie nai 0/3

R sono bun tabere ba ii n desu yo 0/8

Mu yoru yonaka sugi te kara kebabu no teikuauto toka 13/9 = 1.464

Ma okashii na 0/2
doko sun de ta no 0/5

Mu Kantaberii Hooru ni sun de ta no 0/7

Ma doko dak ke 0/3

Mu ano ne Rasseru Sukuea no sugi soba ni Intaakarejji Hooru ga tonton to tsuzui te ru no shit te ru 0/2, 8/18 = 0.444

ano naka no hitotsu 0/6

Ma aa ima Mina chan iru tokoro ka 0/1, 2/6 = 0.333

R sou 0/1

229
Ma  Mina chan iru toko
Mu  aa/anouchi nensei no ko
R   otoko darake no joshiryou to iu hanashi
Ma  nani sore
Mu  sou/otoko darake nano
Ma  nani sore
R   otoko no ko ga ooi ryou nano
Mu  joshiryou na noni otoko ga ippai/2/6 = 0.333
I    nee/Mutsue chan soutou ichou ga joubu na n ja nai/8/14, 5/14 = 0.5
Mu  sou na n desu
R   sugoi kenkou sou na kao shi te ru yo ne
I    (datte)datte sugoi non de mo nanni mo kao ni mo de nai shi/19/15 = 0.667
Ma  kesshoku ii mon ne
Mu  aa/ujji
R   chou ijime rare te ru
Mu  nanse, hachi kiro hutot tat tsuu koto de/2/8 = 0.250
I    demo, hachi kiro hutorut te sugoi yo/4/8 = 0.571
Ma  a/kii te ii
Mu  doko karaha hutoru no hachi kiro hutoru toki/4/8 = 0.500
Mu  kore ga hatachi o koe te kara hutoru basho ga kawaru no desu/15/13 = 0.846
Ma  kao, doko

230
Ma  hora/watashi no hou ga zutto takai desho / o/4, n/8 = 0.142
I  att/hontou / o/4, o/4
Ma  nana juu nai to omou / o/4
watashi ga mina chan to hanashi te mo' mesen kou da mon / n/10 = 0.100
I  att/hontou da / o/4, o/4
mesen ga u da wa / n/5 = 0.150
Ma  watashi no nana juu ichi, jasuto ka son gurai da mon / n/10 = 0.100
I  att/hontou / o/4, o/4
Y  josei no shinchou ga koujou-shi ta nee / o/4
Mu  shi ta nee / o/4
R  (antara) antara, nani yatte / n/4 = 0.150
Ma  Izumi san hosoi kara / o/4
Mu  uun / Izumi san hosoi yo nee / o/4, o/4
I  hoso-sugi de kimochi warui / n/4 = 0.250
Mu  taijuu nan kiro an no / n/6 = 0.400
I  uun / ima nee yon juu hachi kiro gurai kana, c/8 = 0.350
Mu  nana juu de yon juu hachi / c/6 = 0.333
I  uun / o/4
Y  isshou no uchi de yon juu hachi kiro o koe ta koto ga nai n
desho / s/16 = 0.342
I  

aru  \[ \text{aru} \]

gou juu go rokkiro made it ta toki ga at te koukou set no toki

toka\[3\] / 3/1620.

233
N  水が引き始め/お/はな

Tm  お金だよ/お金/はな

N  それが雪か/ても/そ

Tk  色んな光くらい/ても/ほとんどの/でも/また/す

N  而ももっともしか/お/はな

Tk  だから今こそ/ ønska /決してはなな

N  みんながいと/て/と

Tk  名家も今日に至るまで/はな

N  <mat te>  もっとも/はな

Tk  いい/小さい/どこ/はなな

Tm  はい/小さい/どこ/はなな

N  <mat te>  もっとも/はな

Tk  宇治川も今日に至るまで/はな

Tm  はい/小さい/どこ

N  <mat te>  もっとも/はな

Tk  いい/小さい/どこ
Tm でこのか)そこのお金がもとのもだけ/\(1/8 = 0.235\)

お金が人生をかえるわよあはれ/\(1/4 = 0.46\)

ほとんどにせよ

N ono mo tsukat na no/\(1/8\)

Tk un akan ka ne)kinou Yukiko ni kii tara akan ka)kuruma de san ka

hi te sore de ono de{dou na, kou, not}te it te ta kara/\(2/6, 2/4 = 0.085\)

N san kai mo hika nakut tat te ii noni ne/\(1/10\)

Tm sore de are na n desho/\(3/6 = 0.50\)

de sare o mi te i ta/\(2/4 = 0.285\)

Tk mi te ta ja nai kedo/\(1/6\)

Tm onii san te iu ka, goshujin/\(1/6 = 0.166\)

Tk mi te ta no/\(1/4\)

N goshujin/\(1/6\)

Tm <goshujin> dann san ga mi te\(2)shokku de nige ta toki ni\(2)kuruma

ni hika re chat te juutarai mitai na/\(4/4 = 0.19\)

Tk <otou san> otou san mo shin da no/\(1/6\)

Tm otou san juutarai na no/\(1/8 = 0.20\)

N juutarai na n dat te/\(1/5\)

Tk otou san, nande tome nai no yo <nee> jibun no otouto na noni/\(1/4 = 0.383\)

N ee/nande daro/\(1/4, 1/2\)

yappari koresa reru no ga kowakat ta no kana/\(2/3 = 0.46\)

235
Tk
sore de hannin mo jisatsu-shi chat ta n desho/1/9 = 0.125

Tm
sou na no/0/3
nande jisatsu-shi/0/2

Tk
nan ka nan ko yoku wakan nai yo ne/0/3

Tm
jisatsu-suru, kurai nara yan na kya ii noni/4/8 = 0.500

N
datte, kore de, shinsou wa, yami no naka da yo ne/4/5, 4/5 = 0.365

Tk
sore zurui yo ne/0/4

Tm
shiri tai naze/0/3

Tk
{nande} nande toka it te/0/4

N
{naze} naze/0/4

Tm
okane desho u yappari/0/4

okane no kireme wa, en no kireme/2/7 = 0.285

N
a, ne/sou ie ba sa Junko ni it ta/0/5, 0/8, 5/8 = 0.375

Tm
Kimutaku/0/8

N
Kimutaku/0/8

Tk
Kimutaku ga, dou shi ta no, dou shi ta no/1/6, 0/6 = 0.66

Tm
ki te ta yo Rondon ni/2/6 = 0.333

Tk
uso/0/8

N
ima ne Rondon ni/3/6 = 0.750

Tk
{uso uso uso} uso/{nande nande nande nande nande} nande/0/8, 0/8

N
nande, son na, ni, kouhun-sun no/8/6 = 0.500
Tk

Ookusufodo Sutorito no, eigo gakkou ni ite n dat te/

Tk

sugoi mi tai/

Tk

son na ni suki dak ke/

N
demo nihon dat tara aeru chansu wa nai yo ne/

Tk

nai yo ne/

Tk

zettai nai kara ima/

N+Tk

ima/

Tk

nani shite ru no/

Tk

iya/gakkou ki te ru no yo/

N
sou/chan ka ne eikaiwa narai ni ki te ru n dat te/

Tk

nande, minna shitte n no/

Tk

gle sono gakkou ga wakan nai/

N
att/Aya no tomodachi, soko ni itte te/

Tk

att/<sou> sou/

Tk

kurasumeeto ka mo/

N
iya/kurasumeeto/

datte yappari, kojin desho / /

Tk

Att/kojin desu ka/

Tk

yan nai desho sore wa/

N
sore wa, kojin ressun desho /

Tk

sou na no /

237
N de2 sono toki ni <nan ka> mi ta rashii Kimutaku/3/9 =0.375

Tk he/0/4

N gakkou no naka de/0/4

Tm bikkuri/0/4

N da kara Takako, moshi mi tai n dat tara ni juu yo jikan haritsui

tara/16/19, 10=66
Transcription JT3 by Izumi Nakamura 1994

S  karuku/kanamari, keigo tsukawa zu ni /0/4, G/3 /0.400
I  att/sou desu yo ne /0/4
M  sou desu yo nee /0/4
S  nakama-kotoba mitai ni naru nde ne /0/6
I  ee/demo, yappari ima koko de wa hajimete to iu koto to /0/4, G/3 /0.400
M  ee /0/4
I  sorekara (ma) dansei to josei no chigai to /0/4
S  ee /0/4
I  ato, nenrei to /2/3 /0.666
S  sou desho u ne /0/4
I  sou desu /0/4
demo jitsu wa ano watashi wa sensei ni ome ni kakaru no wa,
hajimete ja nai n desu /G/3 /0.400
S  sou desho u ne /0/4

<nan ka> dok ka de ne oai-shi ta you na ki ga /0/4 /0.400
doko deshi tak ke /0/4
<nan ka> seminda ka, dok ka /2/4 /0.500
I  anoo/nihongo kyoushi no yousei kouza yousei kouzat te iu ka
ichii ichi kouza ni kyonen no ni gatsu surai ni sanka-shi masu
<te /0/4, G/3 /0.562

239
aa/sou desu ka / 0/1, 0/3

sono toki ni koushi o shi te rasshat te / 0/35

aa/sou desu ka / 0/1, 0/3

hai/anoo ichi ni koushi na node, sensei ga ya te rasshat
ta no wa ni jikan toka ichi jikan toka ni jikan, toka / 0/3, 0/3, 0/3, 0/3

sou desu yo ne / 0/4

sono koushi na node kedo / 0/3, 0/3, 0/4

aa/un dok ka de ne / 0/1, 0/3, 0/4

sou/are desu yo / 0/1, 0/3

sono koushi na node / 0/1, 0/2

sou are desu ka / 0/1, 0/3

ichi hajime mashi te to / 0/5

mou / 0/1

ichou hajime mashi te to / 0/5

aa/sou desu ka / 0/1, 0/3

sore itsu / 0/4

ni gatsu no / 0/3

aa/jaa mou ichi nen gurai mae / 0/1, 0/5, 0/6 = 0.833

un / ichi nen / honto ni / 0/1, 0/2, 0/2

un / 0/1

ichi nen mae / 0/5
M  choudo ichi nen mae /  Votes 
S  ichi nen mae ne /  Votes 

<nan ka> sono toki nigatteru nan ka don na naityou de ne watashi go are desu ka ne / 2012-09-01 

I  eetto/kekkyoukai bijinesumai ni taisuru koko nihongo kyouiku to iu koto dat ta node de ano dipuroma koosu no setsumei o jare ta n dat ta sare ta n desu kedo / 2012-09-01 

S  a/sou desu ka / 2012-09-01 
I  hai / 2012-09-01 
S  uun/sou desu ne / 2012-09-01 

zenzen mou oboe te mase n / 2012-09-01 

I  ima mo^ dipuroma koosu oshie te rassharu n desu ka / 2012-09-01 
S  (ima) dakara ima mo yat te masu ano sono rongu koosut te iu / 2012-09-01 
I  ee / 2012-09-01 
S  (koko no ano ma) ima koko de yat te n no wa ano nihongo o ne gakubu to koko no eguzamu saabisu de ni kasho de ne nihongo o yat te ru n desu kedo mo ne / 2012-09-01 
I  aa / 2012-09-01 
M  gakubu no hou de mo oshie te rassharu n desu ka / 2012-09-01 
S  gakubu wa (^ma wa) dakara watashi wa ima wa oshie te mase n kara / 2012-09-01 
M  uun / 2012-09-01 

241
I  watashitačhi mo / o/2

M  kedo (ano) ichi gakki desu ka / 2/5 = 0.400

S  eeto / dakara, itsu deshi tak ke / 0/3, 2/5 = 0.200

M  att / mou zennendo desu / o/1, o/3
This thesis started by making a distinction between syntactic complexity and difficulty; complexity refers to the degree of complexity which a given syntactic structure measures whereas syntactic difficulty is relevant to the processing difficulty of a given phrase which arises specifically because of its syntactic properties. The aim of the work reported here is to investigate what aspects of syntax contribute to processing difficulty in Japanese, against the background of the majority of research which assumes that it is all due to just one aspect, namely 'complexity'. In answering this question the research also applies Word Grammar, which seeks to express syntax in terms of direct relationships between words, to Japanese in which the status of word is ambiguous owing to its agglutinative nature.

The recent theories of syntactic complexity reviewed in the thesis are the Maximal Local Nonterminal Count (Frazier & Rayner 1988), the Thematic Interpretation Theory (Gibson 1991, 1995), the Early Immediate Constituents Theory (Hawkins 1994) and the Syntactic Prediction Locality Theory (Gibson 1998). These theories all produce measurements of the computational loads for assigning syntactic structure, all of which are analysed by some version of Phrase-structure Grammar (Chomsky 1981, 1986a; Pollard & Sag 1994; Bresnan 1982). Also they are all based on the assumptions that syntactic processing demands processing resources of working memory and that the more complex syntactic structures are, the more difficult they are to process, so the more processing resources are required.

A variety of theories of processing difficulty such as the ones mentioned above are proposed in order to account for processing overload phenomena, where it is still impossible to arrive at the correct understanding of a sentence even when the appropriate syntactic analysis for it has been made. A well-known example of this type of sentence that has a processing overload effect is one with multiply nested structures as in (9.1).

(9.1) #The frightened child who the old woman who the rescue worker looked for had comforted survived the crash.
    (Gibson, Thomas & Babyonyshev 1995)

It should be remembered that these theories are different in nature from
theories of ambiguity resolution (MacDonald, Parmeuter & Seidenberg 1994, Trueswell et al. 1994), which account for the processing difficulty caused by a reanalysis difference due to a local ambiguity, so they do not make any predictions about the processing overload of the kind in (9.1). An example sentence containing a local ambiguity (a “garden-path” effect) is given below in (9.2).

(9.2) The cotton clothing is made of grows in Mississippi.
(Marcus 1980)

Since the investigation on syntactic difficulty presented in this thesis was not involved in processing difficulty caused by local ambiguity, the theories of ambiguity resolution and the gaps between the two distinct sources of syntactic difficulty were not considered and these issues are left for future research. This thesis was rather concerned with the relative syntactic difficulty in many different types of Japanese sentences which are still processable, so the theories of processing overload were more relevant.

Most models of syntactic parsing, as mentioned above, assume the complexity of a syntactic structure to be the main, or even only, cause for syntactic difficulty, and therefore predict that local processing load is higher at the point the more complex structure is parsed. This has been supported by considerable research such as eye-fixation durations (Frazier & Rayner 1982), self-paced reading and lexical decision times (King & Just 1991). This thesis was not only based on the same assumption, which is however expressed differently in terms of dependency distance - the longer a dependency is, the more difficult it is to process - but also explored other sources of syntactic difficulty such as an effect of word order.¹

Three pilot experiments and one additional experiment in a design of more

¹Other psycholinguistic work supports this view of multiple influences. For example, there is a series of experiments by Bock utilising syntactic priming technique, which show word order in speech is determined by a number of interacting factors (Bock 1982); an animate noun tends to be the subject of a transitive verb (McDonald, Bock & Kelly 1993), conceptually more accessible items tend to appear early in a sentence (Bock & Warren 1985; Bock 1987) and items containing given information tends to be placed before ones with new in a sentence (Bock & Irwin 1980). Since these experiments are essentially involved with syntactic planning in language production, the issues about the relationship between the effect of word order in comprehension and that in production is left for future research.
proper rigor reported in this thesis were designed so as to explore the various aspects of syntax affecting syntactic difficulty. The conclusions that have been drawn from the results of these experiments remain tentative because of the weak control on potential independent variables, improper counterbalance for order of presentation, and insufficient sampling.

The effects on memorability in short-term memory tasks for which some evidence have been found in the experiments are listed below in Table 9.1.

<table>
<thead>
<tr>
<th>Effects</th>
<th>Source</th>
<th>Relevant experiments</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Serial order effect</td>
<td>memory capacity</td>
<td>Experiments 1, 2, 3 &amp; 4</td>
</tr>
<tr>
<td>2. Word-length effect</td>
<td>memory capacity</td>
<td>Experiment 1</td>
</tr>
<tr>
<td>3. Chunking effect</td>
<td>phrasing</td>
<td>Experiment 1</td>
</tr>
<tr>
<td>4. Distance effect</td>
<td>dependency distance</td>
<td>Experiment 2</td>
</tr>
<tr>
<td>5. Chaining effect</td>
<td>dependency distance</td>
<td>Experiment 2</td>
</tr>
<tr>
<td>6. Overload effect</td>
<td>incomplete phrases</td>
<td>Experiment 2</td>
</tr>
<tr>
<td>7. Head effect</td>
<td>phrasing</td>
<td>Experiment 4</td>
</tr>
<tr>
<td>8. Function-word effect</td>
<td>morpho-syntax</td>
<td>Experiment 1</td>
</tr>
<tr>
<td>9. The maximal analysis</td>
<td>morpho-syntax</td>
<td>Experiments 3 &amp; 4</td>
</tr>
<tr>
<td>10. Clausal-order effect</td>
<td>word order</td>
<td>Experiments 3 &amp; 4</td>
</tr>
<tr>
<td>11. Topic-order effect</td>
<td>word order</td>
<td>Experiment 4</td>
</tr>
</tbody>
</table>

1. Serial order effect
All the experiments involved short-term memory span for sequences of differing types or normal sentences, based on the idea that the greater the memory load of the items in a sequence is, the less well they will be recalled.

It was found from the serial position curves of all the experiments that the level of recall was higher at the initial and final sections of the sequences, usually consisting of a few words. The good performance at the beginning and the end of a list is well known, often referred to as the primacy and recency effects, respectively (cf. Rundus 1971).

---

2Regarding Experiment 1, although the utilised sequences were set in a fixed length in terms of mora and those less familiar or hard to pronounce were not included in the sequences, some words in sequences may have formed a chunk, as mentioned, enhancing subjects' performance. In addition the sequences were read out by myself to several groups of subjects rather than presented as recorded materials, which may have resulted in a somewhat inconsistent presentation and should have been avoided.

246
2. Word-length effect
Although the evidence for this effect was not statistically established, it was found in Experiment 1 that the average success rate for the two sequences composed of twelve words of 48 morae long, 58.37%, was better than that for the two sequences composed of twelve words of 60 morae long, 43.33%. This conforms to the word-length effect (Baddeley et al. 1975; Cowan et al. 1992), which accounts for the fact that recall of lists composed of shorter words in terms of time of pronunciation is better than that of lists composed of longer words, since the mora is generally considered as a timing unit (cf. Beckman 1982; Port et al. 1987). This is also consistent with the claim by Caplan, Rochon & Waters (1992) that the main factor in covert rehearsal is not the duration of pronunciation but phonological structure, because the mora is a crucial unit in Japanese phonology (cf. Haraguchi 1991).

3. Chunking effect
In Experiment 1 we found that the average success rate for the three natural sentences, 81.55%, was superior to that for the random sequences, 56.79%, and also the success rate for one natural sentence, nat sent 2, 89.16% was better than that for its scrambled counterpart, 58.83%. This can be attributed to the chunking effect (Miller 1956, 1958; cf. Wingfield & Nolan 1980 for syntactic segmentation of ongoing speech) that is responsible for better recall of (syntactic) chunks and is uncontroversial.

4. Distance effect
Experiment 2 as well as Experiment 3 primarily concerned the relationship between dependency distance as a measure of working memory load and performance on immediate recall tasks. These experiments involved minimal pairs that differed only in terms of dependency distance and which were presented with an interval of a week, between the sessions in which the two sentences were presented, which resulted in “long-term” Hebb effects (Hebb 1961) from week 1 to week 2. It is therefore crucial in conducting this type of experiment to counterbalance the order of presentation.

Despite the small number of examples, Experiment 2, whose minimal pairs contained the sequences with the intervening long relative clauses between the root-verbs and their dependents, tentatively supported the effect of dependency distance that the longer the distance between a parent and its dependent is, the harder they are to process (in this immediate recall tasks,
the harder they are to recall). Interestingly the results were in agreement with the data by Hawkins (2001) and Yamashita (in press) showing a preference for placing a longer phrase before a short one rather than a short-before-long pattern in Japanese speech.

5. Chaining effect & 6. Overload effect
The other syntactic influences we tentatively deduced from Experiment 2 were the “chaining effects” and “overload effects”. The chaining effects are relevant to the enhancement of memorability on “chains” - words linked with dependencies of distance zero - in the initial and rear positions of a sequence. (In the experiment, the subjects’ performance on the chains on these positions was better than expected throughout the chains.)

The “overload effects” refer to the deterioration of memorability on the items that are preceded by words in which there are multiple unintegrated dependencies. This deterioration in performance is understandable because all these unintegrated dependencies must be preserved in working memory, so the working memory should be under great pressure. Utilising sentences with one intervening relative clause, King & Just (1991) and Gibson & Ko (1998) respectively showed in their self-paced word-by-word reading experiments that the longest reading time was recorded on the sentential verbs, whereby the long unintegrated dependencies of subjects were linked with them. Although their finding is not directly related to the overload effect, it seems to be the case that a long dependency or an unintegrated one consumes working memory resources. Since the chaining and overload effects were by-products of the results of Experiment 2, further experiments specifically designed to test these effects must be administered in order to confirm them properly.

7. Head effect
The “Head effect” was derived from the profile of the serial position curve of Experiment 4, which showed that the last word in a phrase, which also happened to be its head in Japanese, was better memorised. A theoretical explanation for this tendency is available if the head word is held in memory as the representative for the entire phrase. The pattern of the curve seems to be relative strong evidence for this interpretation, but the validity of the effect must be statistically confirmed. Moreover it was impossible from the Japanese data to decide which syntactic influence is more relevant to short-
term recall tasks: the last word of a phrase or the phrase's head. These two interpretations will need to be separated in future research. The nature of the head effect will be left for further research.

8. Function-word effect
This thesis provided as one of the key issues a parsing model based on Word Grammar. It is important to implement natural language processing systems into parsing algorithms, as Bresnan (1978) argued that grammar should be compatible with a psychologically realistic parser. Because the parsing algorithm itself has no direct relevance to this work, the one presented in the thesis was very simple-minded, guiding the search for dependencies in such a way that a surface structure is well-formed by allowing the parser to have access to all available data about each word. (For more plausible parsing algorithms based on Word Grammar, refer to Fraser (1985, 1993), Hudson (1989) and Shaumyan (1995)).

However a word-based parsing model requires a clear definition of “word”. Since Japanese is an example of an agglutinative language and also uses no word space in the writing system, it raises a problem for syntactic analysis. Word Grammar, as opposed to some Phrase-structure based approaches to syntax, is compatible with morphology-free syntax (Zwicky 1992: 354), where the internal structure of a word is opaque and irrelevant to syntax. Having analysed what constitutes a word in Japanese according to the principles of morphology-free syntax, we came to a general conclusion that most morphemes are separate words in Japanese. ³ This analysis has the great advantage of explaining why these morphemes follow the words to which they are attached. If the latter are their dependents, then this position fits the very strong head-final tendency of Japanese word order.

To be more precise, in addition to obvious self-standing words like nouns, the following morphemes are treated as individual words:

³From a psychological point of view, it would be interesting to investigate to what degree compounds, phrasal expressions, idioms, and common collocations ought to be treated as a chunk, virtually one psychological word, since they have profound impact on working memory in terms of language processing. In view of the fact that the more frequently they are used, the more likely they become lexical, the lexicalisation of formulaic speech seems to depend on the speaker's own experience and to be difficult to theorise about without psycholinguistic research. (As mentioned before, formulaic speech was avoided in the items of the experiments presented in this thesis.) Hence, although this is a key issue, I shall leave it for future psycholinguistic research.
(i) derivational morphemes that change the word-class of a preceding word, i.e. particles and auxiliary verbs as in (9.3).

(ii) apparently "inflectional" morphemes such as grammatical particles and auxiliary verbs, as shown in (9.4).

(9.3) genki 'healthiness'  
genki na 'healthy'  
noun  
genki ni 'healthily'  
noun  

(9.4) shinu 'die'  
shina nai 'not die'  
verb  
shine ba 'die' (conditional)  
verb  

KEY: cop = copula; neg = negative; ptc = particle

However, derivational morphemes changing the meaning of the word to which they are attached are treated as affixes, so they do not constitute separate words, e.g. (9.5)

(9.5) hi- kokusai- shugi- sha 'anti-internationalist'  
anti- international -ism -ist

As one exception to this, a classifier, e.g. (9.6), was controversially treated as a separate syntactic word on the grounds that its combinations with ten possible digits would lead to the implausible proliferation of lexemes. I shall leave the issues concerning the wordhood of classifiers open.

(9.6) hashi ichi zen 'a pair of chopsticks'  
chopstick one classifier

In sum, Japanese Word Grammar recognises morphemes as words, the exceptions to which are derivational morphemes changing the meaning of the word to which they are attached, and are part of the word.

Experiment 1 tentatively confirmed that the memory load of random sequences with a fixed phonological length increases with the number of
morphemes, from which it follows that function words, particles and auxiliary verbs, contribute to memory load (Function-word effect). This confirmation seems to stand as psycholinguistic evidence, albeit not very strong, for function words in Japanese being separate words. However there was stronger confirmation from the results for the maximal analysis, below.

The random sequence made up of numbers and classifiers presented the confounding result that it was easier to recall than any other random sequence used in the experiment. A possible reason was that subjects were allowed to use the Roman numbers, with which numerals can be written down much faster than with Hiragana letters. I acknowledge this as a flaw in the experimental design. In order to determine whether Japanese function words contribute to working memory load, this type of study must be put to proper examination, which I shall leave for future research.

9. The maximal analysis
Dependency distances for Japanese sentences can be measured in two ways: by analysing morphemes as part of a word (the minimal analysis) and as separate words (the maximal analysis). Experiments 3 & 4 investigated the appropriateness of measuring distance in terms of these two analyses.

As for Experiment 3, in spite of the unreliability of the results due to improper counterbalancing of the order of presentation of the sequences, the results showed some consistency of effects; the results were more consistent with the minimal analysis in terms of dependency distance but not consistent with the results of Experiment 1.

On the other hand, Experiment 4 avoided these design weaknesses. As each minimal pair contained the same word with different clausal order, performance of each individual subject was under the influence of Hebb effects, but the results for the whole were not influenced by Hebb effects thanks to better counterbalancing of the groups. (Another way to counterbalance order of presentation in this type of study using minimal-paired sentences is to present one type of the pair to a half of subjects and the other type to the other half of subjects. This might be preferable in that subjects only engage in different sentences with different words, so the results will not be affected by Hebb effects. But the price paid is the loss of control over effects of inter-subject variation.)
In contrast to Experiment 3, Experiment 4 produced results that were in favour of the maximal analysis in terms of dependency distance, though there was a significant interaction between pair and type (discussed in 11 below). Since the evidence for the appropriateness of measuring dependency distance in terms of the maximal analysis stands as suggestive, I shall leave this issue for further research.

10. Clausal-order effect
We tentatively ascribed the results of Experiment 3, which were opposed to those of Experiment 4, to the effect of the order of clause elements (S/O before A versus A before S/O). We observed better performance on the S/O A sentences than on the A S/O ones in Experiment 3 and also in the two pairs, (4) and (13), in Experiment 4. Assuming that the S/O A order is the default for Japanese, we may also assume that it is easier to process than the A S/O order. From this merely suggestive evidence we can only hypothesise that this effect may be involved in both the experiments, especially in Experiment 3, in addition to the effect of dependency distance. However there is no conclusive psycholinguistic evidence for the effect of this type of clausal order, so I shall leave this issue open.

11. Topic-order effect
After looking even more carefully than usual at the data of Experiment 4, I found the items in the experiment were not strictly counterbalanced in terms of word order and hypothesised that another word-order effect involving topicalisation might be responsible for the interaction noted above between pair and type in the results of the analysis of variance. The hypothesis that AS'V ordered sentences are more difficult to recall than S'AV ordered ones seems to be consistent with the claim by Bock & Irwin (1980) that speakers construct sentences so as to provide given information first before new information, because given information is marked by the topic marker in Japanese by default. However, the evidence for an effect of topicalisation on short-term recall tasks is merely suggestive.

In conclusion, there is some evidence to support the claim that there are several different kinds of syntactic influence on memorability. Most of the evidence is suggestive rather than strong, but sufficient to cast serious doubt on the widely held.
Following the preliminary report of the experiments, this thesis presented a pilot study of spoken English and Japanese texts, whose main concern was to compare the syntactic difficulty of these texts in the two languages merely in terms of mean dependency distance. The analyses were tentative due to the lack of control over topic, setting and speakers of conversation and its diminutive size of sample and also because they were solely word-based, so chunking between words was not given consideration. (In turn it seems very difficult to set a basis for determining a chunk as the appropriate unit of analysis.)

The background to this investigation was the observation that consistent word order (such as in Japanese) tends to increase dependency distance compared with an otherwise similar inconsistent word order (as in English). The question, therefore, was whether actual usage was more difficult (in terms of dependency distance) in Japanese than in English. The results indicated that the degree of syntactic difficulty in speech resulting from dependency distance was similar in English and Japanese. Also we tentatively argued for the two different directions to which grammar may evolve in order to minimise difficulty; English grammar minimises distance by directing dependencies to both sides, whilst Japanese grammar does so by allowing more optionality.

This textual study also seems to suggest that syntax should not only be designed to help the hearer, since the speaker also has needs which may be different from those of the hearer, so Japanese grammar may help the speaker as well as the hearer. As mentioned above, the interesting work on syntactic planning by Bock & Irwin (1980) shows that the speaker composes sentences in such a manner that given information is delivered before new, which seems to reduce processing difficulty in Japanese for both the speaker and the hearer. Clearly much more work needs to be done in this particular area to verify or falsify the close relationship between processing demands (minimisation of dependency distance) and well-formed syntactic structure (optimality). I shall leave the issue open.

All in all this thesis has stressed the inadequacy of other theories which focus on a single measure of syntactic complexity. It is true that considerable research has found evidence for syntactic complexity associated with increased processing difficulty and has suggested that the working memory
system is called on in syntactic processing - assigning the syntactic structure of a sentence. However I have rejected the conventional assumption that syntactic complexity and difficulty are identical, admitting that complexity can be a source of difficulty, and also suggested tentatively that other sources such as word order can contribute to difficulty or ease as well.
References


Verbal Behavior, 14, 575-589.


Craik, F.I.M., & Tulving, E. (1975). Depth of processing and the retention of


Frazier, L. (1985). Syntactic Complexity. In D. Dowty, L. Karttunen & A. Zwicky (Eds.), *Natural Language Processing: Psychological, 259*
Computational and Theoretical Perspectives. Cambridge, UK: Cambridge University Press.


262


263


Pickering, M.J. (1996). *Processing local and unbound dependencies: an eye-
tracking study.


Treiman, R., & Danis, C. (1988). Short-term memory errors for spoken syllables are affected by the linguistic structure of the syllables. *Journal of*


properties and motivations for production. *Text.*


'commodified' goods. Similarly, the increasingly commonplace use of certain types of prestige object over time generally means that their commodity status becomes clearer and their regimes of value more fixed. This slow development from luxury to commodity (and sometimes back again) is a common feature of the way societies respond to new elements in their material culture.

It is helpful to see such transformations in terms of a process of value-articulation (see fig. 2.1). There is a tension inherent in the incorporation of foreign, highly crafted and/or rare objects into existing social contexts (Helms 1988, 1993; Thomas 1991: 35-6). Such objects are potentially very powerful because they carry the prestige associated with restricted areas of knowledge. This knowledge can be geographical in the case of artefacts from far-away places, ideological (including religious knowledge) or technological. Emphasis is often placed on fitting, honourable or knowledgeable employment of this value – the ability of an individual or group to do this in acceptable ways is as much a part of an object's potential prestige value as the mere fact of its possession. In other words, any form of manipulation of such material culture is often highly structured such that only the appropriate people make use of it (and thereby lay claim to its power) at the appropriate time and in the appropriate setting. This structuring is therefore a form of articulation – in a sense the genesis of a social language (Tilley 1991: 17-20) through which the object can be incorporated into the wider ideologies of the society in question (displayed in tabular form in fig. 2.1). This contrasts sharply with many sorts of commodities for which the social framing of their use was less a matter for intense concern because they are already completely a part of the information system. In other words, for these commodities, the process has to some extent already occurred (often in the very distant past) and their value fits within broadly accepted frames. This is often reflected in an accepted exchange value (a price), whereas objects with less articulate value are seen as priceless, unconvertible and/or sacred.