Event-Related Brain Potentials in the Processing of Japanese Wh-Questions

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy
in Linguistics and Cognitive Science

by

Mieko Ueno

Committee in charge:

Professor Robert Kluender, Chair
Professor Victor Ferreira
Professor Shige-Yuki Kuroda
Professor Marta Kutas
Professor John Moore
Professor Maria Polinsky
Professor Yasu-Hiko Tohsaku

2003
The dissertation of Mieko Ueno is approved, and it is acceptable in quality and form for publication on microfilm:

[Signatures]

University of California, San Diego
2003
To all the women who try
TABLE OF CONTENTS

Signature Page ........................................................................................................ iii
Dedication .............................................................................................................. iv
Table of Contents ................................................................................................ v
List of Figures ...................................................................................................... xi
List of Tables .................................................................................................... xv
Acknowledgements .......................................................................................... xvi
Vita ................................................................................................................ xii
Abstract .......................................................................................................... xxiii

1. Introduction .................................................................................................... 1
   1.1. Overview ............................................................................................... 1
   1.1.1. Japanese Wh-Questions ...................................................................... 1
   1.1.2. Research Questions ............................................................................. 4
       1.1.2.1. Question 1: To what extent are syntactically distinct
               languages processed the same way in the brain? ............................ 5
       1.1.2.2. Question 2: How is the scope of wh-in-situ calculated in
               real-time processing? ......................................................................... 10
       1.1.2.3. Question 3: To what extent do neural processing data
               reflect linguistic theory? ........................................................................ 10
       1.1.2.4. Question 4: What is the nature of the relationship (if any)
               among the data obtained from different methodologies? .................. 12
   1.2. Background ............................................................................................. 13
       1.2.1. Theoretical Introduction .................................................................. 13
       1.2.2. Methodological Introduction .............................................................. 17
           1.2.2.1. Acceptability Ratings ................................................................. 18
           1.2.2.2. Self-paced Reading Times ......................................................... 19
           1.2.2.3. Event Related Brain Potentials (ERPs) ....................................... 20
               1.2.2.3.1. ERP Overview ............................................................... 20
4.5.3. Are syntactically distinct languages processed the same way in the brain? .......................................................... 223
  4.5.3.1. Slow Anterior Negative Potentials ........................................ 223
  4.5.3.2. Active Filler Strategy: Incremental Processing ................. 227

4.5.4. To what extent do neural processing data reflect syntactic structure? ........................................................................ 230

4.6. Conclusion ......................................................................................... 232

5. Reading Time Measures of Bi-Clausal Japanese Wh-Questions and Methodological Comparisons .............................................. 234
  5.1. Introduction .................................................................................. 234
    5.1.1. Miyamoto and Takahashi (2001, 2002b) (M&T) .................... 234
    5.1.2. Aim of Present Study ........................................................... 236
    5.1.3. Predictions ............................................................................. 236
  5.2. Norming ....................................................................................... 239
  5.3. Reading Time Study ...................................................................... 245
    5.3.1 Methods .................................................................................. 245
      5.3.1.1 Subjects ........................................................................... 245
      5.3.1.2 Materials ......................................................................... 246
      5.3.1.3 Procedure ......................................................................... 246
      5.3.1.4 Data Analysis .................................................................... 247
    5.3.4. Results and Preliminary Discussion ........................................... 248
      5.3.4.1. Comprehension Task Accuracy ........................................ 248
      5.3.4.2. Self-Paced Reading Times ................................................. 249
      5.3.4.3. Comparison to Miyamoto & Takahashi (2001, 2002b) .... 253
    5.3.5. Discussion: The Present Study and Miyamoto and Takahashi (2001, 2002b) (M&T) .......................................................... 257
  5.4. Methodological Comparisons ..................................................... 260
    5.4.1. Reading Times and Rating Scores ........................................... 263
    5.4.2. ERPs and Rating Scores ......................................................... 263
LIST OF FIGURES

Note: All the ERP figures are shown with low-pass filtering (< 7 Hz), unless otherwise noted.

Figure 1.1 T-model of grammar with four levels of representation ........ 13
Figure 1.2 X-bar template .......................................................... 15
Figure 1.3 D-structure and S-structure representations of What did Calvin bring? .......................................................... 16
Figure 1.4 D-structure and S-structure representations of Calvin-ga nani-o motteka-ndesu-ka (‘What did Calvin bring?’) ....................... 17
Figure 1.5 Overview of how event-related brain potentials (ERPs) are extracted .............................................................. 21
Figure 1.6 Some ERP components .............................................. 23
Figure 1.7 Placement of electrodes following the international 10/20 system ................................................................. 24
Figure 2.1 Flat (non-configurational) structure analysis of scrambling ...... 31
Figure 2.2 Movement (configurational structure) analysis of scrambling ... 32
Figure 2.3 Mean acceptability ratings on experimental materials ............ 44
Figure 2.4 ERPs at midline and parasagittal electrodes to sentence positions between filler and gap of scrambled vs. in-situ conditions ................................................................. 53
Figure 2.5 ERPs at midline and parasagittal electrodes to sentence positions between filler and gap of scrambled demonstrative vs. demonstrative-in-situ questions ........................................... 55
Figure 2.6 ERPs for all electrodes to the pre-gap position of scrambled demonstrative vs. demonstrative-in-situ questions ......................... 58
Figure 2.7 ERPs for all electrodes to the pre-gap position of scrambled wh- vs. wh-in-situ questions ....................................................... 61
Figure 2.8 ERPs for all electrodes to the post-gap verb position of
scrambled demonstrative vs. demonstrative-in-situ questions… 65

Figure 2.9  ERPs for all electrodes to the post-gap verb position of scrambled wh- vs. wh-in-situ questions…………………. 67

Figure 2.10 Phrase structure trees of stimulus sentences under the movement analysis of scrambling by Saito (1985)………… 91

Figure 3.1  Q-operator binding mechanism………………………………………………………… 106

Figure 3.2  Mean acceptability ratings on experimental materials……………… 118

Figure 3.3  ERPs for all electrodes to wh-words and demonstratives……… 126

Figure 3.4  ERPs at a left frontal electrode to four types of pronominal elements……………………………………………………… 129

Figure 3.5  ERPs to words following wh-words up to and including the verb-question particle complex………………… 134

Figure 3.6  ERPs to the final verb-Q position of wh-in-situ vs. yes/no-in-situ-questions…………………………………… 137

Figure 3.7  The F₀ contours of wh-question and yes/no-question……………… 147

Figure 3.8  The F₀ contours of the last six words of questions…………………… 150

Figure 4.1  Mean acceptability ratings of experimental materials……………… 165

Figure 4.2  ERPs to the embedded verb position of embedded clause wh vs. matrix clause wh questions…………………………………… 174

Figure 4.3  ERPs to the embedded verb position of embedded clause wh-vs. yes/no-questions…………………………………… 176

Figure 4.4  ERPs to the matrix verb position of embedded vs. matrix wh-questions………………………………………………………… 179

Figure 4.5  ERPs to the matrix verb position of matrix wh- vs. yes/no-questions………………………………………………………… 181

Figure 4.6  Mean amplitude N400 responses for matrix clause wh vs. yes/no-questions at the matrix verb-Q position for parasagittal electrodes………………………………………………………… 182

Figure 4.7  Mean amplitude N400 responses for matrix clause wh vs.
yes/no-questions at the matrix verb-Q position. Mean amplitudes are given averaged across right vs. left electrodes of the parasagittal and temporal arrays.

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.8</td>
<td>ERPs to the matrix verb position of embedded vs. matrix yes/no-questions.</td>
</tr>
<tr>
<td>4.9</td>
<td>ERPs to all wh- vs. non-wh-modifiers.</td>
</tr>
<tr>
<td>4.10</td>
<td>ERPs to ‘what.kind.of’ vs. adjectival modifiers.</td>
</tr>
<tr>
<td>4.11</td>
<td>ERPs to wh-numeral ‘what.kind.of’ vs. non-wh-numeral ‘2-cup-of’ classifiers.</td>
</tr>
<tr>
<td>4.12</td>
<td>ERPs to the embedded object position of all wh- vs. yes/no-questions.</td>
</tr>
<tr>
<td>4.13</td>
<td>ERPs to the embedded verb position of embedded vs. matrix wh-questions.</td>
</tr>
<tr>
<td>4.14</td>
<td>ERPs to the embedded verb and matrix subject position of embedded vs. matrix wh-questions.</td>
</tr>
<tr>
<td>4.15</td>
<td>ERPs to the from embedded to matrix verb position of embedded vs. matrix wh-questions.</td>
</tr>
<tr>
<td>4.16</td>
<td>ERPs to the embedded verb position of matrix wh- vs. yes/no-questions.</td>
</tr>
<tr>
<td>4.17</td>
<td>ERPs to the embedded verb and matrix subject position of matrix wh- vs. yes/no-questions.</td>
</tr>
<tr>
<td>4.18</td>
<td>ERPs to the region from embedded to matrix verb position of matrix wh- vs. yes/no-questions.</td>
</tr>
<tr>
<td>4.19</td>
<td>The $F_0$ contours of embedded clause wh-questions, matrix clause wh-questions, embedded clause yes/no-questions with unaccented and accented modifiers, and matrix clause yes/no-questions with unaccented and accented modifiers.</td>
</tr>
<tr>
<td>4.20</td>
<td>Relationship between wh-in-situ and Spec, CP.</td>
</tr>
<tr>
<td>5.1</td>
<td>Mean acceptability ratings on experimental materials.</td>
</tr>
</tbody>
</table>
Figure 5.2  Mean reading times per region for the four experimental conditions………………………………………………………….. 250
Figure 5.3  Mean reading times per region for the four experimental conditions (present study)……………………………………………… 254
Figure 5.4  Mean reading times per region for the four experimental conditions (Miyamoto & Takahashi, 2001; 2002b)……………… 254
Figure 5.5  Mean acceptability ratings on four experimental conditions…… 261
Figure 5.6  Mean reading times between embedded and matrix verbs (regions 4-7) of four experimental conditions………………………… 261
Figure 5.7  ERPs between embedded and matrix verbs of four experimental conditions……………………………………… 261
Figure 5.8  ERPs to the embedded verb and matrix subject position of embedded vs. matrix clause yes/no-questions……………… 266
**LIST OF TABLES**

<table>
<thead>
<tr>
<th>Table 2.1</th>
<th>Sample stimuli ........................................................................ 40</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 2.2</td>
<td>Filler .................................................................................. 42</td>
</tr>
<tr>
<td>Table 2.3</td>
<td>Point of comparisons ................................................................ 50</td>
</tr>
<tr>
<td>Table 2.4</td>
<td>Results of overall ANOVAs for the pre-gap region ...................... 57</td>
</tr>
<tr>
<td>Table 2.5</td>
<td>Summary of results ................................................................... 68</td>
</tr>
<tr>
<td>Table 3.1</td>
<td>Sample stimuli ........................................................................ 114</td>
</tr>
<tr>
<td>Table 3.2</td>
<td>Filler ................................................................................... 116</td>
</tr>
<tr>
<td>Table 3.3</td>
<td>Characteristics of four pronoun types collapsed across electrodes and subjects ........................................... 129</td>
</tr>
<tr>
<td>Table 3.4</td>
<td>Mean $F_0$ maxima (in Hz) of relevant words ................................. 149</td>
</tr>
<tr>
<td>Table 4.1</td>
<td>Sample stimuli ........................................................................ 161</td>
</tr>
<tr>
<td>Table 4.2</td>
<td>Filler .................................................................................. 163</td>
</tr>
<tr>
<td>Table 4.3</td>
<td>Mean $F_0$ maxima (in Hz) of relevant words ................................. 215</td>
</tr>
<tr>
<td>Table 5.1</td>
<td>Sample stimuli ........................................................................ 240</td>
</tr>
<tr>
<td>Table 5.2</td>
<td>Filler ................................................................................... 242</td>
</tr>
<tr>
<td>Table 5.3</td>
<td>Mean reading times per region for the four experimental conditions ................................................................. 250</td>
</tr>
<tr>
<td>Table 5.4</td>
<td>Values used for methodological comparisons ............................... 260</td>
</tr>
</tbody>
</table>
ACKNOWLEDGEMENTS

The work in this dissertation was supported by the community and enthusiasm of many researchers in various academic disciplines. I am deeply indebted to each and all of them.

First and foremost, my deepest gratitude goes to my advisor Robert Kluender for his patience, limitless support, and friendship over the years. He has guided me through all the phases of my academic research, and I truly aspire to his vision and ability for higher-level thinking. I would also like to thank my committee members: Masha Polinsky for her valuable advice on many aspects of my academic life; John Moore for his objective eyes on syntax; Yuki Kuroda for always having answers to my Japanese syntax questions; Yasu-Hiko Tohsaku for his assistance with experimental materials and for sharing his perspective as a Japanese linguist; Marta Kutas for her expertise, careful attention to my dissertation draft, and encouragement; and Vic Ferreira for his helpful feedback perspective as a psycholinguist.

I extend my gratitude also to Amalia Arvaniti for her expertise and assistance with prosody analysis; Sharon Rose for helping me through several stages of my graduate career; and Liz Bates for her wisdom and spirit.

I am truly grateful to the people at the Language and Brain Lab, the Computerhelp team, the Center for Research in Language, the Kutas Lab, and the Brain and Cognition Lab, who have helped me tremendously with this project: Wind Cowles for showing me ERP techniques; Matt Walenski for giving me perspective; Ronald Ohst, Sean Perry, and Esmé Cowles for their technical ERP support; Ezra Van
Everbroeck, Dennis Fink, and Marc Silver for their multi-faceted technical support; Robert Buffington and Luis Avila for their help with the reading time paradigm; and Kara Federmeier, Seana Coulson, Laura Kemmer, and Nicole Wicha for sharing with me their wealth of wisdom and techniques in ERP methodology.

I would also like to thank my LIGN 199 research assistants for their assistance at various stages of this research: Lisa Gustafson, Yoko Hahn, Mai Hiragane, Carmel Jequinto, Hitomi Koyama, Kanako Kozasa, Orika Kunimoto, and Junko Suzuki. This work was also made possible with the support of almost one hundred native speakers of Japanese who participated in the experiments or gave me preliminary input on the experimental materials and methodologies. Moreover, special thanks must go to Anne Sumnicht for proofreading the entire volume of this dissertation and Shannon Casey for being my statistics guru.

Many thanks to my classmates, who have kept my school life vivid and delightful: Anne Sumnicht, Robin Thompson, Wind Cowles, Alicia Muñoz, and Gina Taranto for laughter, scrapbooking, and trips to Julian and the Glen Ivy Spa; Nayoung Kwon for her Asian-touch massage and sharing my evening hours at school; Ezra Van Everbroeck for his pointed sense of humor; Michael Hughes, Mimu Tsujimura, Dan Brassil, Jeremy Boyd, Neal Peters, and Shin Fukuda, for keeping me company in the termroom; Michael Klieman for sharing his gourmet cuisine; and Nitya Sethuraman and Shannon Casey for guiding me through the steps towards graduation and the job search.
I would also like to thank Mayumi Mochizuki Mckee, Noriko Kikuchi Knickerbocker, and Hifumi Ito of the Japanese Studies Program for making my Japanese teaching experience joyful, and Judy Sharp and Margaret Marshall of the Theatre and Dance Department for their ballet instruction and keeping me healthy and enchanted during my dissertation years.

I am also indebted to many professors at in the University of Michigan Linguistics department, which is my second home: Michel DeGraff for opening the door to syntax; Pam Beddor for opening the door to experimental work; André Cooper and Rosina Lippi-Green for showing me how to teach; and Karen van Hoek and Andrew Carnie for connecting me to UCSD. I would also like to thank my classmates: Stephanie Lindemann, Jill Dowdy, Miryoung Kim, Moses Hall, Shak, and Judy Dyer for sharing the very beginning of my graduate student life (and Stephanie Lindemann for sharing my conference life afterwards); Eva Juarros, Bill Bahr, and Stephanie Harves for their passion for and inspiration in syntax; Aman Kumar for more discussion over more beer; and Ana Ostermann for sharing her insights in discourse analysis.

I would like to express my sincere gratitude to the researchers in the language processing community for their insightful comments and stimulating discussion, and for motivating me to meet their professional standards: Janet Dean Fodor, Susan Garnsey, Hiroko Hagiwara, Yuki Hirose, Masako Hirotani, Atsu Inoue, Kiwako Itoh, Noriko Iwasaki, Edith Kaan, Yuki Kamide, Nina Kazanina, Kow Kuroda, Reiko Mazuka, Michiko Nakamura, Kentaro Nakatani, JJ Nakayama, Hajime Ono, Colin
Phillips, Doug Roland, Laurie Stowe, Hiroko Yamashita, and Masaya Yoshida. Edson Miyamoto deserves special mention for the numerous email exchanges we have had over the years, which always showed his profound interest in and insights into language processing, and for providing me with vital input to this project.

My appreciation extends to my fellow Japanese friends and superiors (senseis and senpais) who moved forward the course of my life: Michitoshi and Ryoko Soga at Western Michigan University, Toshiro Sako of Aoyama Gakuin Women’s Junior College, Masanao and Natsuko Moriwaki from the University of Michigan, and Tamio Morikawa of Toyota Tsusho Corporation. I would also like to express my gratitude to my teachers who fostered my early intellectual curiosity: from Kami-Itabashi Daini Junior High School, Kinuyo Maekawa for giving me my very first English lesson; Koichi Mizoue for planting the seeds for my love of the study of language; and Kimiko Ikeda for inspiring me in the study of science; and from Tokyo Metropolitan Institute Takehaya Senior High School, Tetsushi Yamada for encouraging me to study abroad; and Tsutomu Urakawa for just loving English.

Finally, I would like to thank my family for letting me follow this path: my father Noboru Ueno, whose unfulfilled dream was to emigrate to Brazil; my mother Sachiko Ueno for managing the home finances to pay for my undergraduate education; my brothers, sister-in-law, and nephew—Akiyoshi, Tomokatsu, Kazue, and Shota Ueno—for taking care of my mother after my father passed away; and my grandmother Mitsuko Takeuchi for always letting me feel her love. Special thanks must also go to my husband Amit Majumdar and his family for their belief in
education. When his late grandfather Hem Chandra Majumdar thought it his duty as a father to give his three daughters the best education possible in the 1950s in Calcutta, who could have imagined that the foundation he laid would help his Japanese granddaughter-in-law attain her doctorate in the United States fifty years later? Amit is one of the most sincere people I have ever known, and has kept the promise he made to me in his Ph.D. dissertation acknowledgements eight years ago: that he would be with me when I receive my Ph.D. The fifteen years we have shared together through the ups and the downs as immigrants at U.S. universities are invaluable. This dissertation literally would not have existed without him.

I gratefully acknowledge financial support for this work from an NIH predoctoral training grant NIHDC00041-08/09 through the Center for Research in Language at UCSD, a dissertation fellowship in the Department of Linguistics at UCSD, a Joseph Naimam graduate fellowship in Japanese Studies at UCSD, Dean’s Social Science research travel funds at UCSD, research and travel grants in the Department of Linguistics at UCSD, and Graduate Enrichment Funds from the Office of Graduate Studies and Research, UCSD. In addition, Robert Kluender has provided financial support for this work from his NIH grant NIHDC02503-01A1.


xx
MA: Cascadilla Press, in press. I was the primary investigator and author of these papers, and the coauthor Robert Kluender directed and supervised the research which forms the basis for these chapters.
VITA

Education

1990  B.A. in Political Science with Class Honors
      University of Michigan, Ann Arbor

1997  M.A. in Linguistics
      University of Michigan, Ann Arbor

2003  Ph.D. in Linguistics and Cognitive Science
      University of California, San Diego

Publications

Japanese /r/ a factor in Japanese listeners' difficulty in perceiving English /l-\r/?
Proceedings of the 16th International Congress on Acoustics and 135th Meeting
Acoustical Society of America (pp. 3000-3001).

Brain and Language, 86, 243-271.

Twenty-Seventh Annual Meeting of the Berkeley Linguistics Society.

Relating grammar and brain. In G. Garding and M. Tsujimura (Eds.), Proceedings
of the Twenty-Second West Coast Conference on Formal Linguistics (pp. 491-

Questions: An ERP Study.

Published Abstracts

Japanese /r/ a factor in Japanese listeners' difficulty in perceiving English /l-\r/?
Journal of the Acoustical Society of America, 103, 3092.


ERP study. Journal of Cognitive Neuroscience, Supplement, 211.
This dissertation examines the processing of Japanese wh-questions using event-related brain potentials (ERPs), reading times, and acceptability ratings. Following brief theoretical and experimental overviews in Chapter 1, Chapter 2 investigates the relationship between a displaced (scrambled) wh-element and its gap by comparing mono-clausal scrambled wh-questions to mono-clausal wh-in-situ questions. Despite the difference in basic word order, the experiment revealed patterns of brain responses similar to those found in English and German, i.e., slow anterior negative potentials between filler and gap, phasic left anterior negativity (LAN) (Kluender & Kutas, 1993; King & Kutas, 1995), and a P600 (Kaan et al., 2000) around the gap. These results suggest there are universal parsing operations for filler-
gap dependencies, which are more compatible with an incremental (Inoue & Fodor, 1995) than head-driven parser model (Pritchett, 1992).

Chapter 3 explores the relationship between a Japanese wh-element and its corresponding Q[uestion]-particle (Cheng, 1991) by comparing mono-clausal wh-questions with structurally equivalent yes/no-questions. Both scrambled and in situ wh-questions elicited right-lateralized anterior negativity (RAN) sentence-finally, relative to yes/no-question counterparts.

Chapter 4 compares bi-clausal matrix and embedded wh-questions to structurally equivalent yes/no-question counterparts. At the embedded clause region, wh-questions elicited sustained AN in comparison to yes/no-questions. Matrix clause wh-questions, in comparison to embedded clause wh-questions, also elicited sustained (R)AN between the embedded and matrix verbs.

These results suggest a reliable neural processing correlate of a wh-element/Q-particle dependency in Japanese, similar to ERP effects seen between wh-fillers and gaps in English, but with a right- rather than left-lateralized distribution. Rather than a local scope calculation process at the verb-Q position, the scope of a Japanese wh-element seems to be licensed by a long-distance incremental linkage with its Q-particle. These results can further be interpreted as the reflection of structural dependencies (i.e., wh-element licensed by COMP).

Chapter 5 examines reading time processing of bi-clausal Japanese wh-questions. The results revealed a continuing slowdown for matrix clause wh-questions, again suggesting a long-distance wh-element/Q-particle linkage.
Chapter 6 concludes that the processing of syntactically distinct languages elicits strikingly similar brain responses, and that there may be some discernible relationship between linguistic theory and brain responses.
Chapter 1: Introduction

1.1 Overview

1.1.1 Japanese Wh-Questions

This dissertation explores the processing of Japanese wh-questions—i.e., questions that contain wh-elements such as ‘what’ and ‘who’—mainly in comparison to the processing of English wh-questions. To investigate this topic, event-related brain potentials (ERP), which are scalp-recorded brainwaves time-locked to an external event (see Section 1.2.2), are used as a major experimental methodology, supplemented by self-paced reading times and acceptability ratings.

Consider English and Japanese yes/no- and wh-questions such as those shown in (1.1) and (1.2). Besides the basic word order difference between the two languages (S(ubject)-V(erb)-O(bject) for English and SOV for Japanese), Japanese wh-questions are different from those in English in several respects. First, Japanese wh-elements do not have to be displaced like English wh-elements. For example, as shown in (1.1b), the wh-object what in English has to be displaced (termed “wh-movement”, see Section 1.2.1) to the beginning of its clause compared to its non-wh counterpart pizza in (1.1a) (the underline in (1.1b) indicates the canonical object position). On the other hand, unlike English, the Japanese wh-object nani-o ‘what-ACC(USATIVE)’ does not have to be displaced (termed “wh-in-situ”, see Section 1.2.1), as shown in (1.2b).
(1.1) English

a. Yes/no-question

Did Calvin bring pizza?

b. Wh-question

What did Calvin bring __?

(1.2) Japanese

a. Yes/no-question

卡尔ビンが ピザを 持ってきたんですか。
Calvin-ga pizza-o mottekita-ndesu-ka.
‘Did Calvin bring pizza?’

b. Wh-question

卡尔ビンが 何を 持ってきたんですか。
Calvin-ga nani-o mottekita-ndesu-ka.
‘What did Calvin bring?’

However, it is possible to displace both non-wh- and wh-objects in front of the subject in Japanese. This phenomenon, called “scrambling” (Saito, 1985), creates a word order similar to that of wh-movement, as shown in (1.3).

(1.3)

ピザを カルビンが 持ってきたんですか。
pizza-o Calvin-ga mottekita-ndesu-ka.
‘Did Calvin bring pizza?’

カルビンが 何を 持ってきたんですか。
Calvin-ga nani-o mottekita-ndesu-ka.
‘What did Calvin bring?’

卡尔ビンが 何を 持ってきたんですか。
Calvin-ga nani-o mottekita-ndesu-ka.
‘What did Calvin bring?’
Second, while scrambling is optional, Japanese wh-words always require a Q(uestion)-particle *ka* or *no* (meaning ‘whether’) at the end of the clause.¹ This Q-particle *ka* determines the interrogative scope of a wh-element, which in essence can be defined as the part of the sentence that is being questioned. As shown in (1.4), the interrogative scope of an English wh-element is transparently mapped onto its position in the sentence. The wh-element *what* can be placed either at the beginning of the embedded clause as in (1.4a), yielding an embedded clause wh-question, or at the beginning of the matrix (or main) clause as in (1.4b), yielding a matrix clause wh-question. On the other hand, as shown in (1.5), the position of an associated Q-particle indicates the scope of a wh-phrase in Japanese. Thus while the wh-elements in the embedded clause wh-question (1.5a) and the matrix clause wh-question (1.5b) remain in situ (not displaced) in the embedded clause, the Q-particle *ka* appears at the end of the interrogative clause, whether embedded (1.5a) or matrix (1.5b). For matrix clause wh-questions like (1.5b), a non-question particle *to* ‘that’ is attached to the embedded verb.

(1.4) English (wh-movement language)

a. Embedded clause wh-question

[Did Hobbes say [what Calvin brought __ ]]? Logical Answer: Yes, he did.

b. Main clause wh-question

[What did Hobbes say [Calvin brought __ ]]? Logical Answer: Pizza.

¹ The exception is when there is a rising intonation in speech, which can be interpreted as a prosodic version of a Q-particle.
(1.5) Japanese (wh-in-situ language)

a. Embedded clause wh-question

[ホッブスは カルビンが 何を 持ってきたか 言ったんですか。]
Hobbes-TOPIC Calvin-NOM what-ACC brought-Q said-POLITE-Q

‘Did Hobbes say what Calvin brought?’

b. Main clause wh-question

[ホッブスは カルビンが 何を持ってきたと 言ったんですか。]
Hobbes-TOPIC Calvin-NOM what-ACC brought-that said-POLITE-Q

‘What did Hobbes say Calvin brought?’

1.1.2 Research Questions

Given the above syntactic features of Japanese wh-questions, this dissertation addresses the following research questions, using the processing of Japanese wh-questions as a case study.

(1.6) Research questions addressed in this dissertation:

a. To what extent are syntactically distinct languages processed the same way in the brain?

b. How is the scope of wh-in-situ calculated in real-time processing?

c. To what extent do neural processing data reflect linguistic theory?

d. What is the nature of the relationship (if any) among the data obtained from different methodologies?

In a broader perspective, this dissertation takes an interdisciplinary approach and aims to bring together theoretical syntax, psycholinguistic processing theories, and cognitive neuroscience. Theoretical syntacticians have been concerned with grammar, which is an abstract representation of language in the mind. They rely on
grammaticality judgments of sentences (testing if a certain sentence sounds grammatical) to construct syntactic theories. On the other hand, cognitive neuroscientists have been traditionally concerned with brain responses to language stimuli, such as words with semantic violations and different probabilities of occurrence, although there has been growing interest in structural phenomena in language processing. In terms of psycholinguistic processing theories, there has been a great deal of research on the processing of English since the 1960s (cf. Fodor, Bever, & Garrett, 1974), yet research on the processing of Japanese, or for that matter of any Asian language, has been in existence for only a little more than a decade (e.g., Mazuka, 1990; Inoue, 1991; Sakamoto, 1991; Yamashita, 1994). In particular, research on the neural processing of Japanese is still in the beginning stages (e.g., Garnsey, Garnsey, Yamashita, Ito, & McClure, 2001; Nakagome et al., 2001; Ueno & Kluender, 2003). Given the above, the goal of this dissertation is to narrow the gap between linguistic and neurological approaches to language processing.

1.1.2.1 Question 1: To what extent are syntactically distinct languages processed the same way in the brain?

Provided that Japanese wh-questions are syntactically distinct from English wh-questions, one might wonder to what extent the neural processing of Japanese wh-questions shows similarities to or differences from the neural processing of English wh-questions. In the psycholinguistic literature, wh-movement languages like English or German are said to involve “filler-gap dependencies”. Consider the English yes/no- and wh-questions shown in (1.1), repeated here as (1.7).
The wh-object what in the wh-question in (1.7b) has to be displaced to the front of the sentence compared to its non-wh counterpart pizza in (1.7a). The displaced wh-element is called a "filler" while the canonical (or ‘not displaced’, ‘in situ’) position is called a “gap”. A filler and its gap are said to be dependent on each other: a filler needs its gap, and a gap requires a filler. As shown in (1.8), a sentence with a filler but no gap (1.8a) and a sentence with a gap but no filler (1.8b) are ungrammatical (ungrammaticality marked by an asterisk).

(1.8)  

a. *What did Calvin bring pizza?  

b. *Did Calvin bring ___?  

It has been said (cf. Fodor, 1978, 1989) that this “filler-gap dependency” occurs because a filler has to be “assigned” to its gap for successful sentence interpretation. At its displaced position, a filler does not give the parser (human sentence processing mechanism) enough information to identify its thematic role (e.g., agent, patient) and grammatical function (e.g., subject, object). These are said to be identified at its gap position following the verb that provides the parser with such information (termed “subcategorization”): 

---

2 There have been claims that it is not the gap but the subcategorizer of the filler (i.e., verb, preposition) that is crucial; a gap is argued to be predicted based on the information supplied by the subcategorizer (Gorrell, 1993;
Previous ERP studies of wh-movement languages like English and German report effects of left anterior negativity (LAN: negative deflection of electrical voltage seen in the left frontal region of the scalp, see Section 1.2.2) between wh-fillers and gaps and around the gap position in response to wh-questions compared to yes/no-questions, which are argued to be an index of working memory load for holding a filler and filling the gap (Kluender and Kutas 1993; Fiebach et al., 2001, among others). Additionally, there have been reports of P600 effects (defined as a positive deflection of electrical voltage around 600 ms poststimulus onset, see Section 1.2.2) at pre-gap (Kaan, Harris, Gibson, & Holcomb, 2000) and post-gap (Phillips, Kazanina, Wong, & Ellis, 2001; Phillips, Kazanina, Garcia-Pedrosa, & Abada, 2003) positions.

The question is whether a language like Japanese would elicit similar effects. Recall from Section 1.1.1 that the difference between Japanese wh-questions and English wh-questions is that (a) Japanese wh-elements do not have to be displaced, but (b) need a Q(uestion)-particle that marks their scope (i.e., which part of the sentence is being questioned).

As for displacement, as discussed in (1.3), it is possible to displace (or “scramble”) both non-wh- and wh-objects in front of the subject in Japanese (phenomenon called “scrambling”) and to create a word order similar to the one created by wh-movement. This might lead to effects that parallel the wh-movement effects found in previous ERP studies of filler-gap dependencies in English. But it has

Gibson & Hickok, 1993), or associating the filler with the subcategorizer itself is argued to be enough (Pickering & Barry, 1991; Pickering, 1993).
to be noted that in a “head-initial language” (see Section 1.2.1) like English, the subcategorizing verb precedes the gap position, and when the filler is assigned to the gap, the parser receives enough information for thematic role and grammatical function assignment. On the other hand, Japanese is a “head-final language” (see Section 1.2.1) in which the verb always appears at the end of the clause. Thus the subcategorizing verb is at the end of the sentence regardless of whether or not the object is displaced. The question then is whether the proposed filler-gap dependency effects also hold for strictly head-final languages like Japanese. More specifically, the question is (a) whether there is/will be any evidence in the ERP record of filler maintenance between a scrambled object and its canonical word order (SOV) position, and if there will be any ERP evidence of gap-filling once that canonical position has been reached, or (b) whether there will be any ERP evidence of the parser waiting until the sentence-final verb to effect a filler-gap association. This also provides a way to test different predictions of an incremental processing model (e.g., full-attachment model, Inoue & Fodor, 1995), in which incoming items are processed before a verbal head, and a head-driven parsing model (Pritchett, 1992), in which incoming words are processed only when the verbal head comes in.

As for the second characteristic of Japanese wh-questions, namely that a Japanese wh-element needs a Q-particle to mark its scope, there have been both a self-paced reading-time study (Miyamoto & Takahashi, 2001, 2002b; Aoshima, Phillips, & Weinberg, in press) and an ERP study (Nakagome et al., 2001) that suggest that the parser expects a Q-particle following a Japanese wh-element. If this is the case, we

3 See footnote 2.
may see some detectable relationship between a Japanese wh-element (either
displaced or in-situ) and its corresponding Q-particle. There are two possibilities for
this relationship. First, the relationship may manifest in the form of a long-distance
dependency between the wh-element and its scope-licensing Q-particle, which may be
similar to the filler-gap dependencies reported in wh-movement languages. As shown
in (1.9), a wh-element is displaced to the left of its gap and marks its scope in English⁴,
while in Japanese a Q-particle is placed to the right of the corresponding wh-element
and marks its scope.

(1.9)

a. English

```
[What did Hobbes say [Calvin brought _]]?
FILLER       GAP
```

b. Japanese

```
[Hobbes-wa [Calvin-ga nani-o mottekita-to itta-ndesu-ka]]?
Hobbes-TOPIC Calvin-NOM what-ACC brought-that said-POLITE-Q
```

‘What did Hobbes say Calvin brought?’

The wh-scope marking system in Japanese may look like the mirror image of the wh-
scope marking system in English, but still creates a dependency and a concomitant
working memory load. In English, the parser has to carry a wh-element without a
thematic role or grammatical function up to the gap position, and this is said to tax
working memory. In Japanese, the parser may have to carry a wh-element without

⁴ Except for a multiple wh-question, such as *Who brought what?*. 
scope information up to the corresponding Q-particle position. This may tax working memory as well, as predicted by Kluender (1998: 268-269):

First, it is true that speakers of wh-in-situ languages are spared the thematic and relational ambiguity of fillers displaced from their logical position in constituent structure, which we have hypothesized results in problems of working memory. But they nonetheless have to cope with the very real problems of scope ambiguity, which one might expect to carry its own processing cost, possibly related to working memory.

However, this is not the only option. Alternatively, we might expect a local interpretative process calculating the scope of wh-in-situ at the corresponding Q-particle position in Japanese. In this process, the parser waits until the Q-particle position that determines the scope of a wh-in-situ, and computes its scope at the Q-particle position as a local process at the simultaneous moment of scope-disambiguation.

1.1.2.2 Question 2: How is the scope of wh-in-situ calculated in real-time processing?

Whatever the answer to Question 1 is would indicate how the scope of Japanese wh-in-situ is calculated in real-time processing. The possibilities include either a long-distance dependency between a wh-element and its scope-marking Q-particle, a local scope calculation process at the Q-particle position, or something unpredicted.

1.1.2.3 Question 3: To what extent do neural processing data reflect linguistic theory?

The question posed here is whether there is any discernible relationship between linguistic claims made with regard to wh-in-situ languages such as Japanese and particular brain responses. Recent advances in neural imaging
technologies allow the testing of linguistic theories using real-time sentence processing. It is now possible to test whether the distinctions made in formal theories of grammar are to be associated with distinct ERP patterns.

In terms of the findings thus far, specific ERP configurations can be associated with fundamental linguistic categories, such as open class vs. closed class words (e.g., Garnsey, 1985; Brown, Hagoort, & Keurs, 1999) and semantics (N400 (see Section 1.2.2): e.g., Kutas & Hillyard, 1980; Federmeier & Kutas, 1999) vs. syntax (P600 (see Section 1.2.2): e.g., Osterhout and Holcomb, 1992; Kaan et al., 2000), although these distinctions may be due to non-linguistic factors (see Section 1.2.2). The question is whether we can see further evidence for a correlation between brain responses and finer theoretical distinctions, in this particular case, related to Japanese wh-questions.

Recall again that the difference between Japanese and English wh-questions is that (a) Japanese wh-elements do not have to be displaced but (b) need a Q-particle that marks their scope. As for displacement, linguistic theory argues that when something is moved (or displaced), the movement operation leaves a trace of the moved element in its original position (base position) and forms a chain between the moved element and its chain (see Section 1.2.1). The question is whether there are brain responses that reflect this chain formation, or at least some form of syntactic dependency. As for wh-scope, there have been several theoretical claims made about Japanese wh-questions, or “wh-in-situ” in general. These claims argue for a mechanism that licenses the scope of wh-in-situ either by a Q-particle (Nishigauchi, 1990; Cheng, 1991) or a Q-operator related to its Q-particle (Baker, 1970; Pesetsky,
1987; Watanabe, 1992; Aoun & Li, 1993; Cole & Herman, 1994). Some linguists (Huang, 1982; Lasnik & Saito, 1984; Nishigauchi, 1990) also argue that wh-in-situ is actually displaced at some level of linguistic representation (LF, see Section 1.2.1) to the same phrase structural position as a Q-operator, and the displaced wh-element itself licenses its scope. These claims linking wh-in-situ with the corresponding Q-particle (or its related Q-operator, or a moved wh in the same phrase structural position as the Q-operator) are consistent with experimental findings (Nakagome et al., 2001; Miyamoto & Takahashi, 2001, 2002a) suggesting that the parser expects a Q-particle following a wh-element. Then it might be reasonable to expect some detectable neural relationship between a Japanese wh-element (either displaced or in-situ) and its corresponding Q-particle. For both characteristics of Japanese wh-questions (displacement and wh-Q), the fundamental question is whether the brain responses we measure reflect the complex relationships (a) between a displaced wh-element and its gap and (b) between an in situ/displaced wh and its scope-marking Q-particle, as proposed in various linguistic theories.

1.1.2.4 Question 4: What is the nature of the relationship (if any) among the data obtained from different methodologies?

In this dissertation, event-related brain potentials (ERPs) are used as the major experimental methodology, as they provide a continuous measure of brain activity during real-time sentence processing. In addition, as supplemental methodologies, acceptability ratings and self-paced reading times are used, in order to obtain norms for the experimental materials and to see whether there are any correlations between particular ERP effects and these measures. The question is whether and how the data
obtained from these three methodologies might be correlated. Some correlations are expected, as all three methodologies are designed to measure certain aspects of sentence processing, but there may be some differences in obtained patterns of results depending on the type of methodology, as in on-line vs. off-line or neural vs. behavioral (see Section 1.2.2).

1.2. Background

This section introduces the basic theoretical and experimental concepts employed in the dissertation.

1.2.1 Theoretical Introduction

The linguistic framework adopted in this dissertation is the Principles and Parameters framework (Chomsky, 1981, 1986), which has been the conventionally used framework for theoretical analyses of wh-in-situ. In this framework, the grammar consists of four levels of representation, namely, D-structure, S-structure, PF and LF. This model of grammar is termed “T-model”, as shown in Figure 1.1.

![Figure 1.1 T-model of grammar with four levels of representation.](image)

---

5 We will also discuss the analyses by the Minimalist Program (Chomsky, 1995), which is a more recent descendant of the Government and Binding theory, as needed.
In this model, D-structure encodes the predicate-argument relations and the thematic properties (e.g., agent, patient) of the sentence, while S-structure reflects the actual ordering of the elements in the surface (visible) string. PF (meaning ‘Phonetic Form’) is the level where phonological and phonetic information is ultimately encoded, whereas LF (‘Logical Form’) is the level where logico-semantic information is represented. Operations of transformations (Move-α) rearrange elements between the different representational levels. The derivation from D-structure to S-structure is called “overt syntax”, in that movement operations result in phonological realizations. The derivation from S-structure to LF is called “covert syntax”: “syntactic” because LF is derived from SS via repeated applications of transformations but “covert” because these transformations do not have phonological realizations. Movement transformations leave behind traces coindexed with their antecedents, forming a chain between the moved antecedent (head of the chain) and the trace (tail of the chain).

In a sentence, words are assumed to be structured into phrases, such as an NP (noun phrase), a VP (verb phrase), an IP (inflectional phrase), and a CP (complementizer phrase) following the template based on X-bar (or X’) theory as shown in Figure 1.2. According to the X-bar theory, each phrase is headed by one phrasal head (X). The head is combined with another phrase as its complement (ZP) to form an X-bar projection (X’). Then the X-bar projection is combined with another phrase (YP) as a specifier (or Spec) (or nothing if there is no specifier) to form the maximal projection of the head X as an XP. Maximal projections (e.g., NP, VP, IP) are also called “constituents” of the sentence.
Figure 1.2  X-bar template.

While the hierarchical order among the projections is assumed to be universal, the linear order of the specifier or the complement in relation to the head is not. A verbal head in English precedes its object complement and results in SVO order, while a verbal head in Japanese follows its object complement and results in SOV order. This is why English is often termed a “head-initial” language, while Japanese is termed a “head-final” language.

To illustrate the above concepts, for instance, the D-structure representation of the sentence *What did Calvin bring?* discussed in (1.1b) is *Calvin did bring what?* as shown in Figure 1.3a, with the wh-object NP *what* “base-generated” in the canonical object position (termed “base position”, identical to what has been referred to as the “gap position” above). When movement transformations apply to the wh-object *what* and in this case also to the auxiliary *did*, the S-structure representation *What did Calvin bring?* as shown in Figure 1.3b, is derived.
Figure 1.3  D-structure and S-structure representations of *What did Calvin bring?*

Notice that by these movement transformations, a chain between *what* and its trace and another chain between *did* and its trace are formed. The heads of these chains, namely, *what* and *did*, are said to “bind” their traces, as they are in a higher position than their traces in the phrase structure and coindexed with their traces (coindexation marked with the subscripts $i$ and $j$).

In contrast to the example in English, the D-structure and S-structure representations of the Japanese counterpart discussed in (1.2b) *Calvin-ga nani-o*

---

6 The internal structure of NPs is omitted to simplify the phrase structure.

7 The treatment of *did* actually varies, and actually it is often assumed not to be present at D-structure. However, here it is treated as present at D-structure for expository purposes.

8 When there is no specifier, the X’-level (in this case V’) may be omitted for simplifying the phrase structure tree.
mottekita-ndesu-ka (‘What did Calvin bring?’) are identical, as shown in Figure 1.4. (Note that a phrase structure tree is read counterclockwise for a head-final language.)

This is why English is called a “wh-movement” language, while Japanese is called a “wh-in-situ” language.

1.2.2 Methodological Introduction

Turning now to the processing side, one of the major goals in language processing research is to find out how the human sentence comprehension system (the parser) operates. A number of experimental methodologies have been employed to address this subject. These methodologies can be classified into a variety of categories, such as off-line vs. on-line or behavioral vs. neural approaches.
“Off-line” methodologies refer to measures that do not rely on the timing of subjects’ responses. They include acceptability rating tasks, sentence recall tasks, and sentence completion tasks. They are said to be useful in assessing subjects’ overall comprehension of stimulus sentences. On the other hand, “on-line” methodologies refer to techniques that measure subjects’ real-time responses as they process sentences. Such methods include word-by-word reading, cross-modal lexical priming, eye-movement observation, and event-related brain potentials. They generally allow moment-by-moment processes to be recorded as they unfold in the time course of sentence processing, although there is some limitation for each methodology.\(^9\)

Alternatively, all non-neural methodologies can be classified into the “behavioral” category as opposed to the “neural” category.

The experiments in this dissertation employ three types of measures that fall into each of the three categories (off-line and behavioral, on-line and behavioral, and on-line and neural): acceptability ratings, self-paced reading times, and event-related brain potentials. The following subsections give brief introductions to these methodologies.

1.2.2.1 Acceptability Ratings

The acceptability rating task is a simple off-line method in which the experimenter asks subjects to rate the acceptability of a given sentence on a certain scale, such as one ranging from one (odd) to five (natural). It has to be noted though

\(^9\) For instance, word-by-word reading can be measured only in terms of reaction times of button press, and cross-modal lexical priming shows activation of separate words that are supposedly related to the sentences subjects are processing. Eye-movement observation may show regressive eye-movement to the words subjects have already read once, and the rapid serial presentation method (see Section 1.2.2.2) used for ERP recording does not or cannot capture this regressive reading.
that some researchers suggest that a scale with a fixed number of categories may obscure finer distinctions in acceptability (Bard, Robertson, & Soece, 1996; Cowart, 1997). They recommend using a method with a more continuous and open-ended scaling system, such as the “magnitude estimation” task (Stevens, 1956), which has traditionally been used to rate human perception in psychophysics. In the present dissertation, a version of this magnitude estimation task was used, in which subjects were instructed to imagine drawing a line whose length indicated how acceptable a given sentence was, and then to choose one of the fixed points on a test card corresponding most closely to the length of their imagined line (Cowart, 1997).

1.2.2.2 Self-paced Reading Times

The self-paced reading task is an on-line reading method, in which subjects are asked to read sentences on a computer monitor at their own pace, while their reading times are recorded. The version used in this dissertation is called a non-cumulative moving-window reading task (Just, Carpenter, & Woolley, 1982). In this method, a stimulus sentence initially appears as a row of dashes, and subjects are instructed to press a button to reveal each subsequent word of the sentence. When subjects press a button, the first word appears and replaces the corresponding dashes. But when subjects press the button again, the first word reverts to dashes, and the second word appears in place of the corresponding dashes. In this way, each subsequent button-press reveals a new word and removes the preceding word (thus the term “moving-window”). The time between two button-presses is recorded as the reading time for each word. This method is said to simulate the eye-fixations for normal reading more
closely than the cumulative reading method, in which words are cumulatively displayed on the monitor, or the rapid serial visual presentation (RSVP) method, in which each word is presented one after the other at the same location on the monitor (Just, Carpenter, & Woolley, 1982; but see Ward & Juola, 1982 for their defense of the RSVP method).

1.2.2.3 Event Related Brain Potentials (ERPs)

1.2.2.3.1 ERP Overview

Event-related brain potentials (ERPs) show electrical activity in the brain that is time-locked to an external event, such as the onset of a stimulus (e.g., a word in a sentence) presented visually or auditorily (cf. Kutas & Van Petten, 1994; Coles & Rugg, 1995; Federmeier & Kutas, 2000). ERPs represent a neural and “electrophysiological” approach to language processing.

As shown in Figure 1.5, ongoing electrical activity in the brain is recorded from electrodes placed on the scalp via an amplifier in the form of the electroencephalogram (EEG), which is said to mostly reflect the summed post-synaptic potentials of the pyramidal cells of the neocortex (cf. Kutas & Dale, 1997). A reference electrode is selected, and the voltage difference between each of the other electrodes and the reference electrode is recorded. The recorded signals are averaged over many trials that belong to the same experimental condition, and across a number of subjects. A brief period (100-400 ms) of pre-stimulus onset is used as the “baseline”, relative to which the voltage deflections of the brainwaves in the post-stimulus region are averaged. This averaging process is deemed to wash out any
activity not related to the stimulus presentation (noise) and extracts only the activity related to the stimulus presentation (signal), which comprises the ERPs. While EEG waveforms are on the order of tens of microvolts, extracted ERP waveforms are usually on the order of microvolts.

Figure 1.5 Overview of how event-related brain potentials (ERPs) are extracted (layout adopted from Hillyard & Kutas, 1983).10

ERPs can offer electrophysiological manifestations of neural activities with resolutions on the order of tens of milliseconds during on-line language comprehension, without the need for an additional task such as a button press. It should be noted, however, that the surface distribution of ERP amplitudes does not directly point to the location of underlying neural generators, i.e., which part of the brain is actually activated. Electrical activity generated in one area can be detected at

10 We thank Laura Kemmer for providing us with the figure.
distant locations. To trace the signal source, one would have to go through neural source modeling techniques, which are still in progress (see Rugg, 1999 for a rather pessimistic view and Kutas, Federmeier, & Sereno, 1999 for a rather optimistic view). Although effects visible at the scalp could be generated by a number of different neural generators (cf. Dale, 1994), distinct scalp distributions strongly suggest some difference in the underlying neural generators.

ERPs consist of positive and negative voltage peaks and are characterized by their polarity (e.g., N for negative and P for positive), poststimulus latency, amplitude, morphology (shape of the waveforms), and distribution over the scalp. Thus the P100 refers to a positivity with a peak latency of 100 ms, while the N400 refers to a negativity with a peak latency of 400 ms (see Figure 1.6; note that negativity is conventionally plotted up). Although these peaks and troughs are often called “components”, what really comprises an ERP component is open to discussion. One classic view (Donchin, Ritter, & McCallum, 1978), however, defines it in terms of a combination of its polarity, latency, distribution, and sensitivity to certain experimental manipulations.
Very roughly speaking, ERP components found around 100-200 ms poststimulus onset are often (but not always) reported to indicate (or “index” in ERP terms) perceptual (vision, audition, attention) processing, 200-300 ms, lexical processing, 300-500 ms semantic processing, and 500-800 ms, structural processing.

Figure 1.7 shows how electrodes are placed over the scalp in the experiments for this dissertation. This figure shows an illustration of a head being viewed from above. The electrode placement follows the international 10/20 system, which indicates the relationship between the location of an electrode and the underlying area of cerebral cortex. The letters on the electrode labels, such as F, T, C, P, and O, indicate lobe location, i.e., frontal, temporal, central (although there is no such thing as a “central lobe”), parietal, and occipital. The letter z refers to an electrode placed on the midline (or “zenith”). Even numbers refer to the right hemisphere and odd numbers to the left hemisphere.

---

11 The “10” and “20” refer to the 10% or 20% interelectrode distance. After taking measurements between certain points on the head (from nasion to inion for the anterior-posterior axis and from left to right post-auricular points for the coronal axis), the electrodes are placed at the points that are 10% and 20% of these distances.
numbers refer to the left hemisphere. The smaller the number, the closer the position is to the midline (cf. Coles & Rugg, 1995).

Figure 1.7 Placement of electrodes following the international 10/20 system.

ERP effects are examined in terms of the relative differences between two (or more) waveforms for (as in ERP components) polarity, latency, amplitude, morphology, and distribution over the scalp. In the example shown in Figure 1.7, each individual plot represents a time course of waveforms for each electrode, ranging from 100 ms prestimulus onset to 900 ms poststimulus onset. Again negativity is conventionally plotted up. The waveforms plotted with a solid line come from
Condition 1 and the waveforms plotted with a dotted line come from Condition 2.

Although we cannot see much difference in the posterior (back) region, we can see that Condition 2 is more negative than Condition 1 in the anterior (frontal), especially right-anterior region, which will be considered a right anterior negativity effect later (see Chapters 3 and 4) in the dissertation. Note that ERPs are relative in three senses: (a) EEG signals consist of the voltage differences relative to the reference electrode, (b) ERPs are derived by averaging EEG signals for the same experimental condition in terms of amplitudes measured relative to a baseline, and (c) an ERP effect to one experimental condition is usually described with respect to another experimental condition. ERP effects can be found in both “phasic” (differences a few 100 msec in duration) and “slow potential” or “slow wave” (differences lasting several seconds) forms. Thus within one experiment, comparisons are often made between various combinations of experimental conditions, at various sentence positions, and within various time windows.

As for the handedness of subjects, about half of left-handers are reported to show left hemisphere dominance in language representation similar to right-handers, but the rest show either reversed asymmetry or a greater degree of bilaterality (cf. Kolb & Whishaw, 1990; Perecman, 1983). In order to minimize inter-subject variability, only right-handed subjects are used for the experiments in the present dissertation.
1.2.2.3.2 Language-related ERP Components

This section lists the major ERP components that have been reported to be language-related, namely, N400, P600, and left anterior negativity (LAN).

The N400 is a negative component with a peak amplitude around 400 ms poststimulus onset, often with a slightly right-lateralized or bilateral posterior distribution. Since its discovery two decades ago (Kutas & Hillyard, 1980), the N400 has been tied to various lexical-semantic context effects, such as semantic congruity (e.g., Kutas & Hillyard, 1980; Kutas & Hillyard, 1983), repetition of words in the experimental context (e.g., Van Petten, Kutas, Kluender, Mitchiner, & McIsaac, 1991), word frequency (e.g., Garnsey, 1985; Van Petten & Kutas, 1991), semantic category (e.g., Federmeier & Kutas, 1999), and prior discourse context (van Berkum, Hagoort, & Brown, 1999). The general principle seems to be that the less expected the word in a given context, the larger the N400 that is elicited in response to it.

The P600 (also called “Late Positive Component (LPC)” or “Syntactic Positive Shift (SPS)”) is a positive component around 600 ms poststimulus onset, usually with a posterior or central distribution. It has been reported for morphosyntactic (e.g., subject-verb agreement) violations (e.g., Hagoort, Brown, & Grusthesen, 1993; Coulson, King, & Kutas, 1998) and for words that are not expected for the preferred reading of the preceding syntactic context (reanalysis or “garden-path” effect) (e.g., Osterhout & Holcomb, 1992; Osterhout, Holcomb, & Swinney, 1994). More recently, it has also been argued that the P600 indexes syntactic integration difficulty in general (Kaan, Harris, Gibson, & Holcomb, 2000), as well as thematic reanalysis (e.g.,
Bornkessel, Schlesewsky, & Friederici, 2003). There have also been claims (e.g., Coulson et al., 1998) that the P600 actually belongs to the family of the P300, which has been generally tied to an “oddball” paradigm and to “context updating” (e.g., Donchin, 1981; Donchin & Coles, 1988)\(^\text{12}\). Possibly related to these claims, there has been a report of the P600 elicited in response to spelling errors (Münte, Heinze, Matzke, Wieringa, & Johannes, 1998).

Left anterior negativity (LAN) includes modulations of negative-going potentials with a frontal maximum that are usually larger over the left than the right hemisphere. There are varieties of this component that come in phasic and slow potential forms. When phasic, LAN effects are often elicited between 300 and 500 ms or 600 ms poststimulus onset, whereas in slow potential forms, LAN effects can continue for several seconds. LAN effects have been reported for phrase structure (e.g., Neville, Nicol, Barss, Forster, & Garrett, 1991) and morphosyntactic violations (e.g., Coulson et al., 1998), as well as in association with increased verbal working memory load in processing filler-gap dependencies (e.g., Kluender & Kutas, 1993a,b; King & Kutas, 1995; Fiebach, Schlesewsky, & Friederici, 2001) and cross-clausal temporal relations (Münte, Schiltz, & Kutas, 1998). In addition, there have been reports of “early LAN (ELAN)”, with an effect found around 200 ms poststimulus onset or even earlier. ELAN effects have been reported for phrase structure violations (e.g., Friederici, Pfeifer, & Hahne, 1993), and have been tied to the ‘first-pass’ parsing processing of structure building (cf. Friederici & Mecklinger, 1996).

\(^{12}\) See Osterhout, McKinnon, Bersick, and Corey (1996) and Osterhout and Hagoort (1999) for counterarguments to this claim.
It has to be noted, however, that these language-related ERP components are unlikely to be solely language-specific. N400 effects may also index the conceptual effort of integrating words into a sentence rather than ‘semantic processing’ per se. LAN effects have been tied to working memory load not only in the language contexts mentioned above but also in non-language contexts (e.g., Ruchkin, Johnson, Canoue, & Ritter, 1990). Likewise, if the P600 really belongs to the P300 family as argued by some researchers, we could say that it indexes a categorization effect in both language and non-language contexts.

1.3 Dissertation Chapter Overview

This dissertation introduces a series of experimental studies that address the above research questions, with the chapters organized as follows.

Chapter 2 examines the processing of Japanese scrambling to test the relationship between a displaced (scrambled) wh-element (or its non-wh counterpart) and its gap. Mono-clausal wh- and yes/no-questions containing scrambled wh- and non-wh- objects are compared with wh- and yes/no-questions containing objects in their canonical order (in situ). Despite the difference in basic word order between English and Japanese, the brain responses to Japanese scrambled wh- and yes/no-questions turn out to be similar to the responses to filler-gap dependencies in English. This replicated result points to the existence of universal parsing operations for filler-gap dependencies. One account for this is parser preference for canonical word order, in that it might be expected that the brain prefers canonical ordering and tries to place displaced elements back into their canonical positions. Another explanation is that
this indicates syntactic complexity due to scrambling, as scrambled elements involve a syntactic chain according to some syntactic analyses.

Chapter 3 begins exploring the relationship between a wh-element and its corresponding Q-particle by investigating the processing of mono-clausal Japanese wh-questions. Both scrambled and in situ wh-questions elicit right-lateralized anterior negativity (RAN) around sentence end relative to their yes/no-question counterparts. However, the question remains whether this effect suggests a long-distance dependency between a wh-element and its Q-particle or a local scope calculation process at the verb-Q complex position.

To answer this question, Chapter 4 examines the processing of bi-clausal Japanese wh-questions with matrix or embedded scope. In the embedded clause region, all wh-questions elicited sustained anterior negativity in comparison to all yes/no-questions. Matrix clause wh-questions, relative to embedded clause wh-questions, elicited (R)AN effects between the embedded and matrix verb positions. Matrix clause wh-questions, in comparison to matrix clause yes/no-questions, also elicited RAN effects at the embedded verb position, with a visible trend continuing through to sentence end. The results suggest a reliable neural processing correlate of the dependency between wh-elements and Q-particles postulated for Japanese, but with a right- rather than left-lateralized distribution. The results also suggest that RAN, as a manifestation of the wh-Q dependency, can be interpreted as the brain’s index of wh-Q dependencies, as discussed in linguistic theories.
Chapter 5 investigates the reading time processing of bi-clausal Japanese wh-questions, in order to explore the possible correlations among the three methodologies used in this dissertation, as well as to test whether the present study will replicate the results of Miyamoto and Takahashi’s reading time study (2001, 2002b), based on whose stimuli the bi-clausal ERP study in this dissertation was designed. The results are basically consistent with the data by Miyamoto and Takahashi. They also reveal that reading time, ERP, and rating measures show a consistent pattern of results with respect to continuing wh-Q dependencies, and the data from these methodologies generally correlate with each other. However, it was also found that each methodology may have particular utility in highlighting some specific aspect of processing.

Chapter 6 concludes the dissertation with a summary of the findings and the answers to the research questions given in (1.6).
Chapter 2: ERP Measures of Japanese Scrambling

2.1 Introduction

2.1.1 Japanese Scrambling

This chapter examines the processing of mono-clausal Japanese scrambling. Japanese is a language with SOV (Subject-Object-Verb) word order: canonically, the subject and then the object(s) precede the verb, which always appears clause-finally. However, subjects and objects can occur in a different order relative to each other. The word order in (2.1a) is canonical, and that in (2.1b) is non-canonical.

(2.1)

a. カルビンが ピザを 持ってきました。
   Calvin-ga pizza-o mottekita.
   Calvin-NOM(INATIVE) pizza-ACC(USATIVE) brought
   ‘Calvin brought pizza’

b. ピザを カルビンが 持ってきました。
   Pizza-o Calvin-ga mottekita.
   pizza-ACC Calvin-NOM brought
   ‘Calvin brought pizza’

Since both word orders are possible, some linguists (Hale, 1980; Farmer, 1980) initially hypothesized that both (2.1a) and (2.1b) have the same flat sentence structure without a verb phrase (VP), as shown in Figure 2.1.

Figure 2.1 Flat (non-configurational) structure analysis of scrambling (Hale, 1980; Farmer, 1980).
Thus, Japanese was initially considered to be a non-configurational, flat structure language that does not have VPs.

However, other linguists (Saito, 1985; Hoji, 1985) proposed that Japanese is actually configurational with a VP level, using the evidence of binding, or the coreference properties of certain pronominals. In the Government and Binding Theory framework (Chomsky, 1981), Saito (1985) proposed that sentences like (2.1b) are created by moving the direct object NP *pizza-o* ‘pizza-ACC’ from its underlying SOV position to the beginning of the clause (specifically, to the specifier of the adjoined S position in the phrase structure configuration in vogue at the time), as shown in Figure 2.2.

![Figure 2.2 Movement (configurational structure) analysis of scrambling (Saito, 1985).](image)

This preposing operation in Japanese is generally referred to as scrambling. Assuming the analysis shown in Figure 2.2, scrambling is one way of creating a so-called “filler-gap” dependency (cf. Fodor, 1978, 1989), in which ‘pizza-ACC’ is the displaced
element referred to as the “filler”, with the canonical object position (marked by $t_i$ in Figure 2.2) representing the site of origin, or “gap”, of this displacement.1

2.1.2 Previous Studies on Filler-Gap Dependencies and Scrambling

There is an abundant literature of behavioral (Holmes & O'Regan, 1981; MacWhinney & Pleh, 1988; King and Just 1991), event-related brain potential (ERP: Kluender & Kutas, 1993a, b; King & Kutas, 1995; Müller, King & Kutas, 1997), positron emission tomography (PET: Stromswold, Caplan, Alpert, & Rauch, 1996; Caplan, Alpert, & Waters, 1998; Caplan, Alpert, & Waters, 1999; Caplan, Alpert, Waters, & Olivieri 2000), and functional magnetic resonance imaging (fMRI: Just, Carpenter, Keller, Eddy, & Thulborn, 1996; Caplan et al., 2002; Cooke et al., 2002), studies that have investigated the role of verbal working memory in the processing of filler-gap dependencies in English. It has long been postulated that associating displaced fillers with their gaps as in (2.2a) involves additional verbal working

---

1 There has been a debate in the literature over the years (Saito, 1985, 1992; Welbelhuth, 1989; Mahajan, 1992) as to whether scrambled elements occupy positions assigned thematic roles (a so-called $A[rgument]$-position) or not (a so-called $Ā$-position). To oversimplify this debate slightly for purposes of exposition, Mahajan and Saito (1992) proposed that clause-internal scrambling can be due to either $A$- or to $Ā$-positions, while clause-external scrambling is of necessity only to $Ā$-positions. This debate is largely irrelevant to our concerns here, however, as we were interested in the processing effects of syntactically displacing constituents (including wh-constituents) from their canonical word order positions in the surface representation of sentences by whatever means.

We will therefore refer to any such displaced element in Japanese as having been “scrambled”, without intending any particular theoretical claims by doing so. Since wh-elements in Japanese generally remain in situ, i.e., in their canonical positions, and do not undergo processes of wh-movement to clause-peripheral positions as do wh-phrases in languages like English, we will also refer to syntactically displaced wh-elements in Japanese as having been “scrambled” in what follows. Scrambled wh-objects were originally included in the experimental design to most closely approximate the manipulations undertaken in previous ERP studies of filler-gap dependencies in English and German via wh-movement. As it turned out, however, our clearest results came from scrambled demonstrative phrases (see Results and Discussion sections).
memory load, inducing an extra processing cost relative to sentences like (2.2b) that do not involve such dependencies.

(2.2) a. What did Calvin bring __?
   b. Did Calvin bring pizza?

   In the ERP literature, this processing cost has been argued to be reflected in a negative voltage deflection with a left frontal maximum between 300 and 500 ms post-word-onset, known as left anterior negativity or “LAN”, which typically first appears immediately after the introduction of a filler in sentence structure. Kluender and Kutas (1993 a, b) examined ERPs at the positions immediately following the filler and the gap of various English wh-questions, and concluded that both storing a filler in working memory and retrieving it for filler-gap assignment are indexed by LAN. King and Kutas (1995) reported related effects in the processing of object and subject relative clauses as in (2.3).

<table>
<thead>
<tr>
<th>Region 1</th>
<th>Region 2</th>
<th>Region 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>(2.3) a. The reporter [who __] harshly attacked [the senator] admitted the error.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. The reporter [who the senator harshly attacked __] admitted the error.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In across sentence averages, King and Kutas found a relatively frontal, bilateral slow negative potential in response to object relative sentences (2.3b) when compared to subject relative sentences (2.3a) in Regions 1 and 3. In addition, as in Kluender and Kutas (1993a), there was a phasic LAN effect immediately following the gap in the object relative clause condition (2.3b) (i.e. at the main verb admitted in Region 3).
This effect was superimposed on and independent of the slow bilateral negative potential ranging across Region 3.²

In terms of German, Kluender and Münte (1998) showed slow left-lateralized anterior negative potentials associated with the verbal working memory load in biclusal object wh-questions relative to biclusal subject wh-questions. Fiebach, Schlesewsky, and Friederici (2001) likewise reported slow negative potentials lateralized to left anterior regions of scalp in response to embedded object wh-questions vs. embedded subject wh-questions as shown in (2.4).

(2.4)

a. Karl fragt sich, WER AM DIENSTAG DEN DOKTOR VERSTÄNDIGT hat.
   ‘Karl wonders who understood the doctor on Tuesday’

b. Karl fragt sich, WEN AM DIENSTAG DER DOKTOR VERSTÄNDIGT hat.
   ‘Karl wonders whom the doctor understood on Tuesday’

They manipulated the number of prepositional phrases (‘on Tuesday’ in (2.4)) and created long (with four PPs) and short (with one PP) questions. By comparing words between a filler and its gap as capitalized in (2.4), they found a sustained frontal negativity for long object questions relative to long subject questions especially over left-anterior electrode sites, although not in short object and subject questions.

Fiebach et al. attributed the effect to the maintenance of the wh-filler in working memory until the syntactic dependency between the dislocated filler and the gap at its

² These effects were essentially replicated in Müller, King, and Kutas (1997) using an auditory version of the King and Kutas (1995) stimulus materials: object relative clause sentences elicited greater anterior negativity starting from the relative clause region (Regions 1 and 2 in King & Kutas) and becoming noticeably larger after the gap (Region 3 in King & Kutas). However, these slow potentials were right- instead of left-lateralized, as had been the case in King and Kutas, in which stimuli were
base position could be established. In addition, Fiebach et al. also reported a P600 effect (see Section 1.2.2 of Chapter 1) at the pre-gap position of wh-object questions and attributed the effect to the establishment of the syntactic dependency between the filler and its gap and the integration of this syntactic chain. This was following a claim made by Kaan et al. based on the P600 effect seen at a pre-gap position of English wh-questions (Kaan, Harris, Gibson, & Holcomb, 2000) that the P600 is an index of syntactic integration processes.

In the present study, sentences with and without scrambled objects were compared to each other, in order to determine if there would be indices of increased working memory load in scrambled sentences similar to those seen in cases of wh-movement in English and German. German is in fact a language that has both wh-movement and scrambling. Rösler, Pechmann, Streb, Röder, and Henninghausen (1998) examined the processing of ditransitive sentences in which subject, indirect object, and direct object (S-IO-DO) were presented in various scrambled orders within verb-final clauses. Subjects responded faster to comprehension questions based on canonical word order S-IO-DO sentences than to questions based on scrambled S-DO-IO sentences. These response time data corroborated findings from an earlier acceptability judgment study using the same stimuli (Pechmann, Uszkoreit, Engelkamp, & Zerbst, 1994), and both measurements corresponded with ERP responses to the same stimuli. The determiner of the scrambled direct object in the S-DO-IO sentences elicited greater anterior negativity, especially over the left side of presented visually. The authors speculated that this was probably due to some mechanism of auditory perception, such as more involvement of the right hemisphere.
the head, when compared to the determiner of the indirect object in canonical S-IO-DO sentences.

By contrast for Japanese, Yamashita (1997) reported that it did not take subjects any longer to read scrambled sentences than it did to read canonically ordered sentences. However, when Mazuka, Itoh, and Kondo (2002) compared canonical order SOV sentences with scrambled OSV sentences, subjects rated the scrambled sentences as more difficult and misleading. This difference in acceptability manifested itself as significant processing costs for scrambled OSV sentences in both self-paced reading and eye-movement studies. Crucially, the subject NP in scrambled sentences like ‘husband-ACC wife-NOM __ awaited’ took longer to read than the direct object NP in canonical word order sentences like ‘husband-NOM wife-ACC awaited.’ Miyamoto and Takahashi (2002a) likewise reported an increased reading time when ditransitive sentences were scrambled within the VP. When the distance between the scrambled filler and its gap was increased by inserting an intervening phrase, the reading time further increased. Mazuka, Itoh, and Kondo (2001) used the same type of stimuli in an ERP study and reported late positivities (P600 effects) in response to both NPs in scrambled OSV (‘husband-ACC wife-NOM __ awaited’) sentences.3

3 Garnsey, Yamashita, Ito, and McClure (2001) also conducted a Japanese ERP study examining sentences with scrambled constituents. Biclausal sentences with the structure S-IO-[DO-S-V]-V (in which the embedded clause direct object is scrambled) were compared to similar sentences with the structure S-IO-[S-DO-V]-V (in which the embedded clause direct object is in canonical position). The embedded clause subjects of both word orders elicited greater bilateral anterior negativity, and the embedded clause subject of the scrambled DO-S-V condition elicited a P600. Garnsey et al. interpreted the bilateral negativity as indexing the opening of a new clause, and the P600 as indexing a garden path effect: participants were led to believe that the scrambled direct object in the DO-S-V condition
2.1.3 Present Study

To sum up, convergent evidence from several ERP studies of filler-gap dependencies suggests that the following ERP components might be relevant in the context of clause-internal scrambling in Japanese: slow negative potentials across the course of a sentence related to holding a filler in working memory, phasic LAN effects between 300 and 500 ms related to retrieving a filler from working memory for purposes of assigning it to a gap, and P600 effects related to integrating a gapped position in the syntactic parse.

In the present study, our main question of interest was whether sentence constituents displaced via scrambling, and thereby deviating from canonical SOV word order, in a strictly head-final language with no syntactic process of wh-movement like Japanese would nevertheless show evidence of being processed by the brain like a filler-gap dependency. In particular, we were interested in seeing if there would be any evidence in the ERP record of filler maintenance between a scrambled sentence constituent and its canonical word order position, and if there would be any ERP evidence of gap-filling once that canonical position was reached, or whether there would instead be ERP evidence that the parser waited until the sentence-final subcategorizing verb before attempting a filler-gap association. The former pattern of results would be predicted by an incremental, full-attachment model (Inoue & Fodor, 1995), in which the parser attempts to assign a filler to a gap at the gap position before the verb, while the latter pattern of results would be predicted by a head-driven parsing.

belonged to the main clause (i.e. canonical word order S-IO-DO) and had to revise this interpretation when the embedded subject came in.
model (Pritchett, 1992), in which arguments and assigned thematic roles are attached into the syntactic parse tree only when the subcategorizing verbal head is processed at sentence end. There was already evidence from behavioral studies of head-final languages like German (e.g., Bader & Lasser, 1994) and Japanese (e.g., Kamide & Mitchell, 1999; Aoshima, Phillips, & Weinberg, in press) falsifying the predictions of a head-driven parser. Thus we suspected that if there were any evidence at all for gap-filling in the ERP record of scrambled sentences, it would be seen early, i.e., at the canonical sentence position of scrambled constituents and before the sentence-final verb position.

2.2 Norming

A paper-and-pencil rating study on a subset of the experimental items was conducted to norm the experimental materials. Stimuli consisted of four conditions of mono-clausal questions, namely, (a) object wh-in-situ questions (hereafter called “wh-in-situ questions”), (b) yes/no-questions with demonstrative objects in-situ (hereafter called “demonstrative-in-situ questions”), (c) scrambled object wh-questions (hereafter called “scrambled wh-questions”), and (d) yes/no-questions with scrambled demonstrative objects (hereafter called “scrambled demonstrative questions”), as shown in Table 2.1 (gap position indicated by underscoring) (see also Appendix 1).
Table 2.1 Sample stimuli

a. Object wh-in-situ questions (wh-in-situ questions)

According to the local newspaper, what did that reckless adventurer finally discover?

b. Yes/no-questions with demonstrative objects in-situ (demonstrative-in-situ questions)

According to the local newspaper, did that reckless adventurer finally discover that?

c. Scrambled object wh-questions (scrambled wh-questions)

According to the local newspaper, what did that reckless adventurer finally discover?

d. Yes/no-questions with scrambled demonstrative objects (scrambled demonstrative questions)

According to the local newspaper, did that reckless adventurer finally discover that?
Filler items consisted of four conditions of interrogative sentences as shown in Table 2.2, namely, (a) ditransitives in canonical word order, (b) ditransitives with accusative-marked demonstrative objects scrambled within the VP preceding the dative objects, (c) ditransitives with accusative-marked demonstrative objects scrambled within S (IP) preceding the nominative subjects, and (d) biclausal embedded clause wh-questions with wh-subjects.
Table 2.2 Filler

<table>
<thead>
<tr>
<th>a. Ditransitive canonical word order</th>
</tr>
</thead>
<tbody>
<tr>
<td>あの出入りの訪問販売員が&lt;br&gt; Ano deiri-no homonhanbaiin-ga→</td>
</tr>
<tr>
<td>その一人暮らしの老女に強引にそれを売りつけたんですか。&lt;br&gt;sono hitorigurashi-norojo-ni goinni sore-o uritsuketa-ndesu-ka.&lt;br&gt;the single-living old-woman-DAT aggressively that-ACC force-sold- POL- Q</td>
</tr>
<tr>
<td>‘Did the door-to-door salesman aggressively push that (product) on the lonely old woman?’</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>b. Ditransitive VP scrambled demonstrative objects</th>
</tr>
</thead>
<tbody>
<tr>
<td>あの出入りの訪問販売員が&lt;br&gt; Ano deiri-no homonhanbaiin-ga→</td>
</tr>
<tr>
<td>それをその一人暮らしの老女に強引に売りつけたんですか。&lt;br&gt;sore-o sono hitorigurashi-norojo-ni goinni __ uritsuketa-ndesu-ka.&lt;br&gt;that-ACC the single-living old-woman-DAT aggressively force-sold- POL- Q</td>
</tr>
<tr>
<td>‘Did the door-to-door salesman aggressively push that (product) on the lonely old woman?’</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>c. Ditransitive IP scrambled demonstrative objects</th>
</tr>
</thead>
<tbody>
<tr>
<td>それをあの出入りの訪問販売員が&lt;br&gt; Sore-o ano deiri-no homonhanbaiin-ga→</td>
</tr>
<tr>
<td>その一人暮らしの老女に強引に売りつけたんですか。&lt;br&gt;sono hitorigurashi-norojo-ni goinni __ uritsuketa-ndesu-ka.&lt;br&gt;the single-living old-woman-DAT aggressively force-sold- POL- Q</td>
</tr>
<tr>
<td>‘Did the door-to-door salesman aggressively push that (product) on the lonely old woman?’</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>d. Biclausal embedded clause subject wh-question</th>
</tr>
</thead>
<tbody>
<tr>
<td>あの料理学校の講師は誰が華々しく&lt;br&gt; Ano ryorigakko-no koshi-wa dare-ga hanabanashiku→</td>
</tr>
<tr>
<td>その本場仕込みの広東料理を披露したか話したんですか。&lt;br&gt;sono honbajikomi-no kantonryori-o hiroshita ka hanashita-ndesu-ka.&lt;br&gt;the authentic Cantonese cooking-ACC demonstrated Q said- POL- Q</td>
</tr>
<tr>
<td>‘Did the cooking school instructor say who showed off the authentic Cantonese cooking?’</td>
</tr>
</tbody>
</table>
Twenty sets (a subpart of the 200 sets used in the ERP experiment discussed below) of experimental sentences were placed in a Latin square design to create four parallel lists such that no one subject saw more than one sentence from each set. Twenty filler items (subpart of the 200 sets used in the ERP experiment discussed below) were added to each list, and then each list was pseudorandomized. 20 native speakers of Japanese (different participants than those in the present experiment, but the same as the participants for the reading time study discussed in Chapter 5) were asked to rate each item on a scale from 1 (odd) to 5 (natural), with ‘natural’ meaning ‘sounds natural to a native speaker, appropriate in ordinary relaxed conversation’. However, as a version of the magnitude estimation task (Stevens, 1956) that allows a more continuous and open-ended scaling system than a five-point scale, subjects were instructed to imagine drawing a line whose length indicated how acceptable a given sentence was, and then to choose one of the fixed points (on a 1-to-5 scale) on a scantron that corresponded most closely to the length of their imagined line (Cowart, 1997, see Chapter 1 Section 1.2.2.1).
The results are shown in Figure 2.3. The mean acceptability ratings on the four experimental conditions ranged between 2.7 and 3.0 [wh-in-situ: 2.8 (sd 1.25); demonstrative-in-situ: 3.0 (sd 1.29); scrambled wh: 2.7 (sd 1.35); scrambled demonstrative: 2.8 (sd 1.27)]. Repeated measures ANOVAs were run with condition as a within-group factor, and either subject (F1) or item (F2) as a random factor. There was no statistically significant difference among the four conditions [omnibus ANOVA: F1(3, 19) = 2.16, p = .093; F2(3, 19) = 1.65, p = .178; no significantly different pairs with Tukey HSD posthoc comparisons]. Even when specific pairs were compared separately, neither the comparison between demonstrative-in-situ and scrambled demonstrative questions [F1(1, 19) = 2.23, p = .152; F2(1, 19) = 2.77, p = .113], wh-in-situ and scrambled wh-questions [F1(1, 19) = .45, p = .508; F2(1, 19) =
wh-in-situ and demonstrative-in-situ questions $[F_1(1, 19) = 2.08, p = .631]$, nor scrambled wh- and demonstrative-questions
$[F_1(1, 19) = .38, p = .543; F_2(1, 19) = .43, p = .518]$ showed a significant difference.

2.3 Methods

2.3.1 Subjects

20 (11 female) native speakers of Japanese between 19-29 years of age (mean: 25) who had been outside Japan for less than 2 years were included in the study. Subjects had normal or corrected-to-normal vision, were right-handed (5 reported left-handed family members), and had no neurological or reading disorders. Subjects were reimbursed for their time.

2.3.2 Materials

Stimuli consisted of four conditions of mono-clausal questions, namely, (a) wh-in-situ questions, (b) demonstrative-in-situ questions, and (c) scrambled wh-questions, and (d) scrambled demonstrative-questions, as shown in Table 2.1 (see also Appendix 1). 200 sets of sentences containing these four conditions were constructed. Fillers consisted of four conditions of interrogative sentences, namely, (a) ditransitives in canonical word order, (b) ditransitives with accusative-marked demonstrative objects scrambled within the VP preceding the dative objects, (c) ditransitives with accusative-marked demonstrative objects scrambled within S (IP) preceding the nominative subjects, and (d) embedded clause wh-questions with wh-subjects as shown in Table 2.2. The in-situ subject wh 
*dare-ga* ‘who-NOM’ was used to prevent subjects from always expecting wh-words with object-marking (*nani-o* ‘what-ACC’ or
Ditransitive sentences were used to lead subjects to expect dative objects as well. Different types of scrambling constructions were used so that subjects would not expect scrambled constituents in any particular sentence position.

The 200 sets of experimental sentences were placed in a Latin square design to create four parallel lists of 200 experimental sentences such that no one subject saw more than one sentence from each set. The 200 filler questions were added to each list, and then each list was pseudorandomized and divided into 20 sets of 20 sentences each.

2.3.3 Procedure

Subjects were run in two sessions lasting about 2.5 hours each. Subjects were seated in a reclining chair facing a computer monitor in a sound-attenuated room and wore an elastic cap mounted with tin electrodes. An illuminated rectangular border appeared uninterruptedly in the middle of the screen during presentation of stimulus sentences for purposes of fixation. Stimuli were presented in Japanese characters basically one bunsetsu at a time with 650 ms duration and 650 ms⁴ stimulus onset asynchrony. A bunsetsu consists of one free morpheme (lexical word or pronoun) and the bound morpheme/s associated with it (particles modifying the noun/verb such as the object and subject markers), and will be referred to as a “word” hereafter. The interstimulus interval between sentences was 3000 ms, and subjects were given as much rest as they wished between sets of sentences.

---

⁴ The presentation rate of 650 ms per bunsetsu is longer than the standard 500 ms per word for English, yet was decided on as the best rate after consulting three native speakers of Japanese. Similarly, Garnsey et al. (2001) used 700 ms per bunsetsu in their Japanese ERP study.
In order to maintain subjects’ attention, comprehension questions were inserted in the stimuli. Every five sentences on average but semi-randomly (at least three and at most seven comprehension questions were inserted in a set of 20 sentences), subjects were asked to answer a comprehension question regarding the immediately preceding sentence. Sentence (2.5) gives an example.

(2.5) Comprehension Question

a. Stimulus:

根据当地报纸，那个无法无天的冒险家最后发现了什么?

b. Question:

说话人问“根据当地报纸”，“那个无法无天的冒险家”最后发现了什么?

Comprehension questions appeared 1000 ms following the offset of the sentence-final word, and remained on the screen until subjects responded to them by pressing one of two buttons held in their hands. The next sentence began 3000 ms after the button

---

5 The response hand was not counterbalanced to control for dominance, yet this is not expected to yield a problem, as the accuracy but not the speed of the response was recorded.
press. Before beginning with the first experimental set, subjects were given a practice set of 20 sentences.

2.3.4 Electrophysiological Recording

The electroencephalogram (EEG) was recorded from 19 positions, including all standard positions of the international 10/20 system, using tin electrodes mounted in an elastic cap. Reference electrodes were positioned on the two mastoid processes, and the EEG was algebraically referenced off-line to the mean of the activity at these two electrodes. To detect blinks and lateral eye-movements for later correction, additional electrodes were placed beneath the right eye and at the outer canthi of the two eyes. Impedances were kept below 5KΩ. The EEG was amplified with a bandpass of 0.01 to 100 Hz, digitized at 250 Hz, and stored on hard disk for off-line analysis. Data with excessive blinks were corrected using a spatial filter algorithm (Dale, 1994), and a band-pass filter set from 0.2 to 15 Hz was used on all the data prior to running analyses to reduce high frequency noise.

2.3.5 Data Analysis

Measurements were taken of single-word averages for phasic effects and of two-word and four-word averages for longer-lasting effects. Single-word averages consisted of 1024 ms epochs including a 100 ms prestimulus baseline; two-word averages consisted of 2048 ms epochs (2 x 650 ms stimulus onset asynchrony [SOA] plus a 100 ms prestimulus baseline); four-word averages consisted of 3072 ms epochs (4 X 650 ms SOA plus a 400 ms prestimulus baseline), unless otherwise noted.
In order to see the effects of scrambling, overall ANOVAs treating movement (in-situ vs. scrambled) and type (demonstrative vs. wh) as factors were conducted. In addition to the two condition factors, the ANOVAs included additional topographical factors. The statistical analyses were done separately on midline (Fz, Cz, and Pz), parasagittal (Fp1/2, F3/4, C3/4, P3/4, O1/2), and temporal (F7/8, T3/4, T5/6) electrodes. Midline analyses consisted of three-way ANOVAs with three within-group factors, including movement, type, and three levels of anterior/posterior sites. Parasagittal analyses consisted of four-way ANOVAs with four within-group factors, including movement, type, two levels of hemisphere, and five levels of anterior/posterior sites. Temporal analyses consisted of four-way ANOVAs with four within-group factors, including movement, type, two levels of hemisphere, and three levels of anterior/posterior sites. Pairwise comparisons were undertaken between in-situ vs. scrambled demonstratives and between in-situ vs. scrambled wh in the face of significant interactions between movement and type in some electrode array/s for a particular sentence position. In such cases, additional ANOVAs were run on the three electrode arrays, treating movement, hemisphere, and anteriority as factors. An alpha level of .05 was used for all statistical tests, with a p-value of .10 considered marginally significant. Following a modified Bonferroni procedure that lets up to [number of conditions -1] contrasts to maintain the same alpha level (Keppel, 1982), an alpha level of .05 was also used for planned pairwise comparisons. The Huynh-Feldt correction for lack of sphericity was applied whenever applicable. Original degrees of freedom are reported with the corrected probability level.
2.4 Results

The mean correct response rate to comprehension questions across subjects was 91% (range: 80%-97%). Thus no subject’s data were excluded from the ERP analyses based on poor comprehension. The blink rejection rate (subsequently corrected by the blink-correction algorithm) was approximately 13%. The rate of artifact rejection for other reasons was approximately 17%, somewhat high probably due to the length of the stimulus sentences.

Table 2.3 shows the English gloss of the sample stimuli. Recall from the introduction that we planned to investigate the effects of holding a filler and of filling a gap.

Table 2.3  Point of comparisons

<table>
<thead>
<tr>
<th></th>
<th>pre-gap region</th>
<th>post-gap region</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. The local newspaper-to according</td>
<td>that reckless adventurer-N finally discovered-Q.</td>
<td>that/what-A</td>
</tr>
<tr>
<td>b. The local newspaper-to according</td>
<td>that reckless adventurer-N finally discovered-Q.</td>
<td>that/what-A</td>
</tr>
<tr>
<td></td>
<td>FILLER</td>
<td>GAP</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>between filler and gap</td>
<td></td>
</tr>
</tbody>
</table>

‘According to the local newspaper, did that reckless adventurer finally discover that/what?’

More specifically, to observe effects of filler storage on the ERP record, we planned to investigate the sentence positions between the displaced object ‘that/what-ACC’ and
the presumed gap before the verb, i.e. ‘that reckless adventurer finally’, as shown in Table 2.3. As for gap-filling effects, there were two possible sentence positions to check for such effects, namely, before and after the posited gap position. For example, in a question with a scrambled object like the one in Table 2.3b, the gap would presumably be between ‘finally’ and ‘discovered-Q’, and comparisons were made at both positions. However, the parser might also try to dispose of the filler as soon as possible (Frazier & Clifton, 1989), and since a sentence like (2.6a) (which places the demonstrative object right after the subject, omitting the adverbial ‘finally’) is also possible, comparisons were also made at ‘adventurer-NOM’.

(2.6)

a. Canonical word order

The local newspaper-to according

that reckless adventurer-NOM that-ACC discovered-Q.

‘According to the local newspaper, did that reckless adventurer discover that?’

b. Scrambled word order

The local newspaper-to according

that-ACC that reckless adventurer-NOM (__) discovered-Q.

‘According to the local newspaper, did that reckless adventurer discover that?’

Therefore, ‘adventurer-NOM finally’ was treated as the pre-gap region, while ‘discovered-Q’ was treated as the post-gap region and examined separately.
2.4.1 Effects of Holding a Filler: Between Filler and Gap

This section examines the effects of filler storage by comparing the ERPs to sentence positions between filler and gap, i.e. ‘that reckless adventurer-NOM finally’ in Table 2.3. Visual inspection of the waveforms elicited by this four-word string indicated that scrambled demonstrative and wh-questions elicited slow potential effects of increased bilateral anterior negativity (AN) in comparison to their in-situ counterparts (Figure 2.4). In addition, a distinct positivity was observed near the end of the first word in the string (‘that’) for scrambled demonstrative questions (Figure 2.5), but not for scrambled wh-questions.

First, to confirm the observation of slow potentials, overall ANOVAs were run on mean amplitude measurements between 950 and 2600 ms, covering the second half of ‘reckless’ and all of ‘adventurer-NOM finally’\(^6\). In this latency window, there was a marginal main effect of movement in parasagittal electrodes \([F(1,19) = 3.99, p = .06]\) and a significant interaction between movement and anteriority in parasagittal \([F(4,76) = 3.59, p = .038]\) and temporal \([F(2,38) = 6.18, p = .005]\) electrode arrays. These indicated that the ERPs to the scrambled conditions were more negative than those to the in-situ conditions at anterior electrodes (see Figure 2.4).

\(^6\) Although the subject NP-adverb complex between filler and gap was preceded by words with roughly equivalent ERPs in three out of four conditions (i.e. ‘according’ in the in-situ conditions, and ‘that-ACC’ in the scrambled demonstrative condition), and therefore could have served as a suitable prestimulus baseline, the scrambled wh-objects ‘what-ACC’ in the scrambled wh-condition elicited a larger P200 (see Chapter 3 for details). Since a pre-stimulus baseline of 400 ms would have included this difference between conditions in the P200 region, measurements were taken with a 100 ms post-stimulus baseline instead, which allowed reliable alignment of the post-stimulus onset ERPs.
それ/何を その 命知らずの冒険家が とうとう 見つけたんですか。

Scrambled: ...that/what-A that reckless adventurer-N finally discovered-Q.

その 命知らずの冒険家が とうとう 見つけたんですか。

In-Situ: ...that reckless adventurer-N finally that/what-A discovered-Q.

Figure 2.4 ERPs (n=20) at midline and parasagittal electrodes (Fp1/2, F3/4, C3/4, P3/4, O1/2, Fz, Cz, Pz) to sentence positions between filler and gap (‘that reckless adventurer-NOM finally’ in Tables 2.1 and 2.3) of scrambled vs. in-situ conditions, collapsed across demonstrative and wh conditions, and relative to a 100 ms poststimulus baseline.
Second, to test the effect of late positivity at the first word of the string (‘that’), separate overall ANOVAs were run in the latency window of 500 and 800 ms. Supporting the observation of a late positivity in response to scrambled demonstrative questions, there was a marginal interaction between movement and type in the midline array [F(1, 19) = 4.09, p = .057], as well as a marginal movement x type x anteriority interaction in the parasagittal array [F(4, 76) = 2.46, p = .10]. Due to the statistical interactions involving type, pairwise comparisons were made between in-situ vs. scrambled demonstrative questions, and between in-situ vs. scrambled wh-questions. When scrambled demonstrative questions were compared to in-situ demonstrative questions, there was a significant main effect of movement in the midline [F(1, 19) = 11.24, p = .003] and parasagittal [F(1, 19) = 4.59, p = .045] arrays, supporting the visual impression that scrambled demonstratives were more positive in this latency window (see Figure 2.5). There was no statistically reliable difference in the scrambled wh vs. wh-in-situ comparison.

---

7 There were also marginally significant differences in the P200 region (150-250 ms) at midline electrodes [F(1, 19) = 4.02, p = .059]. However, we believe that this cannot be the sole cause of the P600 effect in response to this word of the sentence. First, there were also significant P600 effects, but no accompanying significant P200 effects, in the parasagittal array in response to the same word. Second, at Fz in the midline, the ERPs to the two conditions diverge in the P200 region, realign in the N400 region, and then diverge again in the P600 region. In addition, there was an effect of scrambling on mean amplitudes at Fz that just missed significance in the P600 region (500-800 ms) [t(19) = 1.68, p = .054], whereas there was no significant or marginal effect of scrambling in the P200 region (150-250 msec) [t(19) = 1.19, p = .124].
Figure 2.5. ERPs (n=20) at midline and parasagittal electrodes (Fp1/2, F3/4, C3/4, P3/4, O1/2, Fz, Cz, Pz) to sentence positions between filler and gap (‘that reckless adventurer-NOM finally’ in Tables 2.1 and 2.3) of scrambled demonstrative vs. demonstrative-in-situ questions, relative to a 400 ms prestimulus baseline.
2.4.2 Effects of Filling a Gap

2.4.2.1 Pre-Gap Region

This section examines the ERPs to the pre-gap region, ‘adventurer-NOM finally’ in Table 2.3, for possible gap-filling effects. Visual inspection of the waveforms suggested that compared to their in-situ counterparts, scrambled conditions (demonstrative and wh) elicited a left frontal negativity roughly around 300-600 ms poststimulus onset of both ‘adventurer-NOM’ and ‘finally’. In addition, there was a positivity in response to scrambled demonstrative questions around 500-800 ms poststimulus onset of both ‘adventurer-NOM’ and ‘finally’. Overall ANOVAs showed complex interactions which corroborated these observations (see Table 2.4). Since these comparisons yielded complex interactions involving type (demonstrative vs. wh), we proceeded to examine pairwise comparisons between in-situ vs. scrambled demonstrative questions, and between in-situ vs. scrambled wh-questions.
Table 2.4 Results of overall ANOVAs for the pre-gap region

<table>
<thead>
<tr>
<th>Movement</th>
<th>'Adventurer-NOM'</th>
<th>'Finally'</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LAN: 300-600 ms</td>
<td>P600: 500-800 ms</td>
</tr>
<tr>
<td>Movement</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Movement x Type</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Movement x Hemisphere</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Movement x Anteriority</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Movement x Type x Hemisphere</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Movement x Type x Anteriority</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Movement x Hemisphere x Anteriority</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Movement x Type x Hemisphere x Anteriority</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*p < .10.  **p < .05.

Visual inspection of the waveforms suggested that scrambled demonstrative in comparison to demonstrative-in-situ questions elicited greater left frontal negativity between 300 and 600 ms poststimulus onset of both ‘adventurer-NOM’ and ‘finally’, as well as a positivity near the end of both ‘adventurer-NOM’ and ‘finally’ (see Figure 2.6).
Figure 2.6 ERPs (n=20) for all electrodes to the pre-gap position ('adventurer-NOM finally' in Tables 2.1 and 2.3) of scrambled demonstrative vs. demonstrative-in-situ questions.
To test the observed LAN effects in response to scrambled demonstrative questions, ANOVAs were performed in the 300-600 ms region of ‘adventurer-NOM’ (latency window: 300-600 ms) and in the 300-600 ms region of ‘finally’ (latency window: 950-1250 ms). At ‘adventurer-NOM’, there were significant [midline: $F(2, 38)=5.29, p = .0298$; parasagittal: $F(4, 76)=3.79, p = .03$] as well as marginal [temporal: $F(2, 38)=3.11, p = .06$] interactions between movement and anteriority, indicating that scrambled demonstrative questions were more negative at anterior regions compared to demonstrative-in-situ questions. At ‘finally’, there was a marginal interaction between movement and hemisphere at temporal electrodes [$F(1, 19)=4.05, p = .059$]: the ERPs to scrambled demonstrative questions were more negative over the left hemisphere but more positive over the right hemisphere compared to demonstrative-in-situ questions. Additionally, there was a marginal movement x hemisphere x anteriority interaction in the temporal [$F(2, 38)=3.12, p = .056$] array, because scrambled demonstrative compared to demonstrative-in-situ questions were more positive over the right hemisphere but more negative over the front and more positive over the back of the left hemisphere.

The observation of a late positivity (P600) in the scrambled demonstrative condition was for the most part supported by ANOVAs run in both the 500-800 ms region of ‘adventurer-NOM’ (latency: 500-800 ms) and in the 500-800 ms region of ‘finally’ (latency: 1150-1450 ms). For ‘adventurer-NOM’, there was a marginal main

---

8The movement x anteriority interaction at midline electrodes was probably due to P600 effects rather than LAN effects. As seen in Figure 2.4, the midline region shows clear positivity at posterior sites starting around 400 ms poststimulus onset of ‘adventurer-NOM’.
effect of movement in the scrambled condition at midline electrodes $[F(1, 19)=3.82, p = .065]$. For ‘finally’, there was a significant interaction between movement and hemisphere at temporal electrodes $[F(1, 19)=11.74, p = .003]$, because scrambled demonstrative questions showed greater late positivity especially over the right hemisphere in response to the pre-gap region. In addition to the above, at ‘finally’, there was a significant movement x hemisphere x anteriority interaction at temporal electrodes $[F(2, 38)=3.61, p = .037]$, due to the fact that scrambled demonstrative relative to demonstrative-in-situ questions were consistently more positive over the right hemisphere, but more negative over the front and more positive over the back of the left hemisphere.

Visual inspection of the waveforms showed an effect of left anterior negativity in response to sentences with scrambled wh-questions between 300-600 ms poststimulus onset of ‘adventurer-NOM’ and ‘finally’ (see Figure 2.7). Unlike the demonstrative comparison, there was no apparent late positivity effect (see Figure 2.7).
Figure 2.7 ERPs (n=20) for all electrodes to the pre-gap position (‘adventurer-NOM finally’ in Tables 2.1 and 2.3) of scrambled wh- vs. wh-in-situ questions.
To test the LAN effect, ANOVAs were run in the 300-600 ms region of ‘adventurer-N’ (latency window: 300-600 ms). The ANOVAs revealed no significant movement-related effect, except for a movement x hemisphere x anteriority interaction in the temporal array [F(2, 38) = 7.7, p = .002], which appeared to be due to a frontal negativity and a posterior positivity in response to scrambled wh-questions over the right hemisphere. Separate ANOVAs were run in the 300-600 ms region of ‘finally’ (latency window: 950-1250 ms). There was a significant [temporal: F(2, 38)=4.76, p = .016] as well as marginal [parasagittal: F(4, 76)=2.61, p = .09] interaction between movement and anteriority, indicating that scrambled wh in comparison to wh-in-situ conditions were more negative over the front than over the back of the head. In addition, there was a marginal movement x hemisphere x anteriority interaction at temporal electrodes [F(2, 38)=2.94, p = .066], because the ERPs to scrambled wh compared to wh-in-situ conditions were more negative over left anterior regions but more positive over right posterior regions.

In summary, overall ANOVAs showed complex interactions among the factors of movement, type, hemisphere, and anteriority in response to the pre-gap region. When pairwise comparisons were examined, LAN effects were elicited in both scrambled vs. in-situ demonstrative and scrambled vs. wh-in-situ comparisons. Although LAN effects were visually apparent at both words for both comparisons, the effect at the subject noun position (‘adventurer-NOM’) for the wh comparison was not statistically supported. P600 effects were elicited at both the subject noun and adverbial positions for the scrambled vs. in-situ demonstrative comparison, whereas
there was no apparent late positivity effect for the scrambled vs. wh-in-situ comparison.

### 2.4.2.2 Post-Gap Region

In order to test if there was any further gap-filling effect at the post-gap position in an SOV language like Japanese, single-word averages were made of ERPs to the post-gap region, ‘discovered-Q in Table 2.3. Visual inspection of the waveforms suggested an anterior negativity in response to both scrambled demonstrative and scrambled wh conditions starting around 300 ms poststimulus onset and continuing to the end of the epoch, although the impression was stronger for scrambled demonstrative questions. Overall ANOVAs performed in the latency window of 300-900 ms supported the observation. Corroborating the observed anterior negativity, there was a significant main effect of movement in the temporal array \[ F(1,19) = 9.53, p = .006 \], as well a significant interaction between movement and anteriority in the parasagittal array \[ F(4,76) = 3.68, p = .033 \]. Reflecting the observed difference between scrambled demonstrative and wh-questions, there was a significant main effect of type in temporal electrodes \[ F(1,19) = 6.63, p = .019 \], as well as a significant \[ midline: F(1,19) = 5.09, p = .036 \] or marginal \[ temporal: F(1,19) = 3.36, p = .082 \] interaction between movement and type. Because of the statistical interactions involving type (demonstrative vs. wh), pairwise comparisons were made between scrambled demonstrative vs. demonstrative-in-situ questions and between scrambled wh- vs. and wh-in-situ questions.
The anterior negativity observed in response to scrambled demonstrative questions was tested by ANOVAs run in the latency window of 300 to 900 ms\(^9\), and yielded a significant main effect of movement in all electrode arrays \(\textit{midline}: F(1, 19)=8.30, p = .01; \textit{parasagittal}: F(1, 19)=9.62, p = .006; \textit{temporal}: F(1, 19)=21.94, p = .001\]. In addition, in the parasagittal array there was a marginal movement x anteriority interaction \(F(4, 76)=2.75, p = .074\]. This indicated that scrambled demonstrative questions tended to elicit greater negativity over anterior regions than demonstrative-in-situ questions at the sentence-final verb position (Figure 2.8).

\(^9\) 300-900 ms is longer than the typical latency window for LAN effects (300-500 or 300-600 ms), yet was chosen since there was no word following the sentence-final verb. To be consistent with the standard window, separate ANOVAs were also run in the latency window of 300-600 ms. In the 300-600 ms window, there was a significant \(\textit{temporal}: F(1, 19)=14.23, p < .001\] and marginal \(\textit{midline}: F(1, 19)=3.31, p = .085\] main effect of movement, indicating that scrambled demonstrative questions elicited greater negativity at the final verb.
Figure 2.8 ERPs (n=20) for all electrodes to the post-gap verb position (‘discovered-Q’ in Tables 2.1 and 2.3) of scrambled demonstrative vs. demonstrative-in-situ questions.
Visual inspection, for the wh-comparison also, suggested an anterior negativity starting around 300 ms poststimulus onset, especially over the left hemisphere, in response to scrambled wh-questions. The observation of an anterior negativity was marginally supported by ANOVAs done in the latency window of 300 to 900 ms\(^\text{10}\). There was a marginal interaction between movement and anteriority at parasagittal electrodes \([F(4, 76)=2.39, p = .105]\), indicating that scrambled wh-questions elicited more negativity at anterior sites compared to wh-in-situ questions at the sentence-final verb (see Figure 2.9).

In summary, both scrambled demonstrative questions as well as scrambled wh-questions elicited greater anterior negativity at the verb position relative to their in-situ counterparts, although only very marginally in the wh-comparison.

---

\(^{10}\) Again, separate ANOVAs were run in a latency window of 300-600 ms, the standard latency window for LAN effects. In the 300-600 ms window, there was a marginal interaction between movement and anteriority at parasagittal \([F(4, 76)=2.60, p = .072]\) electrodes, indicating that scrambled wh-questions elicited greater negativity over the front of the head, but greater positivity over the back of the head relative to wh-in-situ questions.
Figure 2.9 ERPs (n=20) for all electrodes to the post-gap verb position ('discovered-Q' in Tables 2.1 and 2.3) of scrambled wh- vs. wh-in-situ questions.
2.5 General Discussion

The results of our ERP study of filler-gap dependencies in Japanese were remarkably similar to the results of previous ERP studies of filler-gap dependencies in English and German (see Table 2.5).

Table 2.5 Summary of results

<table>
<thead>
<tr>
<th></th>
<th>Scrambled demonstrative vs. Demonstrative-in-situ</th>
<th>Scrambled wh vs. Wh-in-situ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between filler and gap 'that reckless adventurer-N finally'</td>
<td>bilateral anterior negativity (slow potentials)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>P600 (phasic)</td>
<td>--</td>
</tr>
<tr>
<td>Pre-gap (?) region 'adventurer-N finally'</td>
<td>P600 (phasic)</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>LAN (phasic)</td>
<td>LAN (phasic)</td>
</tr>
<tr>
<td>Post-gap region 'discovered-Q'</td>
<td>anterior negativity (phasic?)</td>
<td>anterior negativity (phasic?)</td>
</tr>
</tbody>
</table>

Japanese questions containing scrambled object fillers elicited bilateral slow anterior negative potentials between filler and gap, P600 effects immediately following fillers and immediately preceding gaps (scrambled demonstrative condition only), phasic LAN effects in the gap region, and bilateral anterior negativity (AN) at the sentence-
final verb position. As will be discussed in greater detail below, we assume that the slow negative potentials indexed storage of the filler in working memory, the P600 effects syntactic integration of the filler-gap dependency, the phasic LAN effects the retrieval of the filler from working memory for purposes of gap assignment, and the AN effects at the final verb a continuing effect of working memory.

The overwhelming familiarity of these results is striking in the face of major differences between Japanese and English in typology (strictly head-final vs. head-initial languages), syntactic operation (scrambling via adjunction [or A-movement] vs. wh-[Á-]movement via substitution), and apparent processing requirements. Due to the head-final nature of Japanese, sentence processing in this language entails the need to accommodate massive amounts of syntactic ambiguity on a regular basis, while the processing of a head-initial language like English requires only occasional tolerance for syntactic ambiguity (cf. Inoue & Fodor, 1995). In the remainder of this section we discuss the nature of our results and their individual idiosyncrasies, compare them to the results of other related studies, and finally address their larger implications for sentence processing.

2.5.1 Slow Potentials Between Filler and Gap (Figure 2.4)

When the subject NP and following adverbial in the scrambled O S __ V configuration were compared to the same four words in the canonical S O V configuration, the former showed greater bilateral, slow potential anterior negativity. The most ready interpretation of this effect was that the parser was holding an extra (object) NP in the O S __ V configuration at the points of comparison, causing an extra
verbal working memory load. These slow potential effects in response to questions
with scrambled constituents were consistent with previous reports in the literature that
such effects are caused by the necessity of holding a filler in working memory pending
its assignment to a gap. This has been demonstrated both in English and in German in
response to processes of wh-movement (i.e. wh-question and relative clause
formation: King & Kutas 1995; Müller et al., 1997; Münte, Schwirtz, Wieringa,
Matzke, & Johannes, 1997; Kluender & Münte, 1998; Fiebach et al., 2001), and for
German alone in response to clause-internal scrambling (Rösler et al., 1998).

The distribution of this effect has been inconsistent across studies, however. In
the present study and in the King and Kutas study of English relative clauses, the
effect had a symmetric anterior maximum. In Münte et al. (1997), which involved two
separate replications of King and Kutas using English and German materials, the
effect exhibited a symmetric centroparietal maximum. A left anterior maximum was
reported in two independent studies of German wh-object vs. wh-subject questions
(Kluender & Münte, 1998; Fiebach et al., 2001) and in the Rösler et al. study of
clause-internal scrambling in German, while Müller et al., using auditory presentation
to replicate King and Kutas in English, reported a right anterior maximum for the slow
potential effect. Given this diversity in distribution of the slow negative potential
effect across studies of filler-gap dependencies, we are not in a position to sort out the
functional significance of its variability, but believe that our results are nonetheless
fully compatible with those of previous studies.
A possible objection to our interpretation of the slow potential effect in our data is that since Japanese is a pro-drop language, i.e. it freely allows null pronominal elements, participants might have interpreted the accusative-marked NP object in the OSV configuration as following a null subject (pro O...V), rather than as a displaced object scrambled from its canonical position. Under such an interpretation, the nominative-marked NP, which was the point of comparison (O S ___V), could have been interpreted as the subject of an embedded clause instead (pro O [S...] V). On this interpretation, the slow anterior negative potentials could have been caused by the need to open a new clause (see Garnsey et al., 2001, fn.3). Note however that the negative slow potential difference between scrambled and in situ conditions began already at the adjective (‘reckless’) of the subject NP (‘that reckless adventurer-NOM’), i.e. before the nominative case-marking on the final head noun of the subject NP had been registered, and thus before the need to open a new clause even arose. Since determiners and modifiers are not case-marked in Japanese, at the adjective position (‘reckless’) it would not have been entirely clear that this NP would be marked nominative. Moreover, even if the nature of our stimulus materials led subjects to expect nominative-marked subjects following initial accusative-marked direct objects, Uehara and Bradley (1998) showed in their sentence completion study that Japanese speakers overwhelmingly prefer to interpret a nominative-marked NP preceded by an accusative-marked NP (O S....) as the overt subject of the highest clause. In other words, Japanese speakers will opt for scrambled word order when the alternative is an empty subject in canonical position in the highest clause, followed by
an overt subject in an embedded clause \((pro \ O [S\ldots] \ V)\). Therefore, it seems reasonable to assume that participants in the present experiment anticipated the scrambled structure when presented with an initial accusative-marked NP, resulting in an increase in working memory load.

A further objection to the verbal working memory interpretation of the slow anterior negative potential could be based on the suggestion that our scrambled conditions were somewhat unnatural. In a corpus study of written Japanese texts (Yamashita, 2002), more than 95% of the scrambled sentences in the corpus involved heavy syntactic elements like subordinate clauses and referential NPs. There were no scrambled wh-words in the corpus, but in general it seems clear that relatively light elements like demonstratives and wh-words are rarely scrambled in written Japanese. Thus the slow anterior negative potential effect could have been merely a response to the unusual nature of our scrambled stimulus materials. However, the acceptability judgment ratings of the experimental materials (see Section 2.2) seems to militate against such an interpretation. Recall that demonstratives-in-situ and wh-in-situ questions received ratings of 3.0 and 2.8, fairly close to those assigned to scrambled versions of the same sentences (2.8 and 2.7 respectively), and that there was no statistically reliable difference in acceptability between the scrambled and in-situ versions. From these acceptability ratings, we can infer that the fact that light rather than heavy sentence constituents were scrambled in our design did not cause major problems that obscured the processing. Given the similarity of the slow potential effects in this study to those in other ERP studies of filler-gap dependencies, we will
assume in what follows that they index differences in verbal working memory load across conditions.

This interpretation has the additional advantage of accounting for the discrepancy in results between our ERP study of Japanese scrambling and that of Mazuka et al. (2001), who reported no slow anterior negative potential effects between filler and gap (or anywhere else) in their comparison of OSV and SOV structures. We would like to suggest that this discrepancy may be due to a simple difference in length of the stimulus materials. The materials used by Mazuka et al. consisted of three-word sentences (‘husband-ACC wife-NOM ___ awaited’), while those in the present study consisted of sentences with ten words. Fiebach et al. (2001) showed that German wh-object fillers separated from their gaps by eight words (presented in five phrases) elicited more sustained left-lateralized anterior negativity across sentence positions than wh-object fillers separated from their gaps by only four words (presented in two phrases). While the shorter condition also elicited transient left anterior negativity at the first phrase following the filler (relative to the same phrase in a wh-subject question of equivalent length), this effect disappeared at the second intervening phrase, which immediately preceded the gap position. In a similar fashion, we intentionally lengthened the region between the scrambled object-filler and its gap in our stimulus materials to include four words, the subject NP (‘that reckless adventurer-NOM’) and the adverb (‘finally’). This manipulation appears to have effectively exacerbated the verbal working memory demands of our scrambled sentences, which in turn elicited the slow potential effects. Mazuka et al., on the other hand, seem to
have intentionally kept their stimulus materials as short as possible, inserting no intervening material whatsoever between the scrambled object filler and the subject position preceding the gap, in effect keeping the working memory demands of their stimulus materials to a bare minimum. This interpretation further predicts that if Mazuka et al. had inserted intervening material between the scrambled object filler and its gap in their design, comparisons made at such intervening sentence positions would have also showed slow negative potential differences. A report of an increased reading time with a longer distance between a scrambled filler and its gap (Miyamoto & Takahashi, 2002a) is indicative of such a prediction.

2.5.2 P600 Effects in Response to Scrambled Demonstrative Fillers and Their Gaps (Figures 2.5 and 2.6)

In scrambled demonstrative questions, the word immediately following the scrambled demonstrative object (i.e. the first word of the subject NP) elicited a marked P600 effect relative to the same sentence position in the in-situ demonstrative condition (Figure 2.5). While this is to our knowledge the first reported instance of a P600 tied to the appearance of a filler, Mazuka, et al. (2001) also reported a P600 to the scrambled direct object NP (‘husband-ACC’) of their O S __V (‘husband-ACC wife-NOM __ awaited’) sentences. We are not sure why the P600 effect would be contemporaneous with the appearance of the filler itself in Mazuka et al., but delayed by one word in our data, but this may have had something to do with the fact that the scrambled direct objects in Mazuka et al.’s design were full referential noun phrases, while ours were instead out-of-the-blue, closed-class (and therefore light) anaphoric demonstratives.
Additionally, we observed an effect of late positivity both at the last word of the subject NP (i.e. the head noun, ‘adventurer-NOM’) and at the adverb (‘finally’) of the scrambled demonstrative condition (Figure 2.6). As discussed at greater length in the next section, we were unable to predict the location of the gap in our scrambled sentences with 100% accuracy. For this reason, we cannot say with absolute certainty whether both of these sentence positions constituted the pre-gap region, or whether the head noun of the subject NP constituted the pre-gap region, and the adverbial the post-gap region. However, this is not of major consequence in this case, since P600 related to gap-filling have been elicited in response to the processing of both the pre-gap (Kaan et al., 2000) and the post-gap region (Phillips, Kazanina, Wong, & Ellis, 2001; Phillips, Kazanina, Garcia-Pedrosa, & Abada, 2003) of object wh-questions in English. In addition, Fiebach et al. (2001) reported P600 effects (superimposed on a slow left anterior negative potential, much like in the present study) at the subject position immediately preceding the gap of verb-final German wh-object questions. Most closely related to the present study, Mazuka et al. (2001) reported P600 effects not only in response to the sentence-initial scrambled object NP (‘husband-ACC’) of their experimental sentences (‘husband-ACC wife-NOM __ awaited’), as discussed in the preceding section, but also to the immediately adjacent subject NP preceding the gap (‘wife-NOM __ ’).

There are two arguments for interpreting these P600 effects as an index of syntactic integration costs rather than verbal working memory costs, as first proposed by Kaan et al. (2000). First, the entry of a filler into working memory seems to be
reliably indexed by the onset of a slow negative potential, rather than a P600 effect, as demonstrated in the many ERP studies of filler-gap dependencies cited in the previous section. Second, and more crucially, recall that the filler and its gap were separated from each other in our data by at least three words (‘that reckless adventurer-NOM [__] finally __’), while in Mazuka et al. (2001), the filler and pre-gap positions were immediately adjacent to each other (‘husband-ACC wife-NOM __ awaited’). In the previous section, we attributed the lack of a slow negative potential effect in the Mazuka et al. study to this difference in the length of the stimulus materials. Note, however, that despite these differences in length, the P600s elicited by scrambled fillers and their gaps were quite similar across the two studies.

Likewise, Fiebach et al. (2001) reported that the subject NP preceding the gap elicited a P600 effect in both their long and their short wh-object conditions. P600 effects seen immediately following (or, in Mazuka et al.’s data, simultaneous with) a filler and immediately preceding a gap thus seem to be independent of filler-gap distance, and hence of verbal working memory demands. The most parsimonious conclusion, then, based on the results of previous ERP studies as well as the present study, is that P600 effects in these contexts must be related to the necessity of incorporating a filler-gap dependency into the syntactic parse, rather than an index of verbal working memory demands per se.

Why we did not elicit similar effects in our scrambled wh-condition is not entirely clear, but the upside of this is that our study and that of Mazuka et al. (2001) were consistent in the elicitation of P600 effects in response to scrambled NP fillers
and their gaps. Note further that so far, P600s in response to (scrambled) fillers have been elicited only in Japanese sentence processing; prior to this, P600 effects had been observed only around gap positions, and only in English and German (Kaan et al., 2000; Fiebach et al., 2001; Phillips et al., 2001, Phillips et al., 2003). Whether this is a reliable cross-linguistic difference is something that must be determined by further research.

2.5.3 (L)AN Effects in the Gap Regions (Figures 2.6 through 2.9)

2.5.3.1 Pre-gap or Post-gap?

There were two sentence positions relevant for potential gap-filling effects, namely before and after the posited gap position. However, as mentioned in the previous section, given the relatively free word order of Japanese, it was difficult to pinpoint the exact location of the gap in our experimental sentences. Examples of our in-situ and scrambled demonstrative conditions are repeated for convenience in (2.7a) and (2.7b), respectively.
(2.7)
a. Demonstrative-in-situ question

The local newspaper-to according that reckless adventurer-NOM finally that-ACC discovered-Q.

‘According to the local newspaper, did that reckless adventurer finally discover that?’

b. scrambled demonstrative question

GAP FILLING?

The local newspaper-to according that-ACC that reckless adventurer-NOM (__) finally__discovered-Q.

‘According to the local newspaper, did that reckless adventurer finally discover that?’

As for scrambled demonstrative questions (2.7b), the position of the gap was ambiguous: it could have either preceded or followed the adverb. As discussed in the Results section, an active filler strategy (Frazier & Clifton, 1989) approach to gap-filling predicts that the parser would have tried to dispose of the filler as quickly as possible, and the earliest possible place to posit an object gap would have been at the head noun of the subject NP (‘adventurer-NOM’). However, there would have been nothing to prevent the parser from confirming this assignment at the following adverbial (‘finally’). In a cross-modal priming study, Nicol (1993) reported filler reactivation at two separate sentence positions: the subcategorizing verb, as proposed by the direct association hypothesis (Pickering & Barry, 1991; Pickering, 1993), and again at the purported gap position later in the string. It thus seems plausible that there would have been multiple attempts at gap-filling in response to our stimuli as
well. And indeed, there were P600 effects and LAN effects alike at both the head noun of the subject NP and again at the adverb.

The effects at the sentence-final verb appeared to be different in character. First, there was no P600 in the ERPs to the verb. This was the pattern of results in Mazuka et al. (2001) as well: although both of the first two words in their three-word scrambled OSV sentences elicited P600 effects, there were no significant effects at all at the final verb. Second, while the LAN effects in response to the head noun of the subject NP and to the adverb in the present study were clearly left-lateralized, the effects of anterior negativity to the final verb were not. Therefore, while we are ultimately unsure how to divide three sentence positions (head noun of the subject NP, adverb, and sentence-final verb) into two well-defined regions, one preceding the gap and one following it, in what follows we will treat the LAN effects at the head noun of the subject NP and the adverb (‘adventurer-NOM finally’) separately from the effects of anterior negativity at the final verb (‘discovered-Q’), and consider them as qualitatively different.

2.5.3.2 Issues on Pre/post-gap LAN

Summarizing the results in the gap region, scrambled demonstrative questions elicited LAN effects (which were followed by P600 effects, as discussed in the previous section) at both the subject noun and adverbial positions, while scrambled wh-questions elicited LAN effects (but no subsequent P600 effect) at the adverbial position only. At the final verb, both scrambled conditions elicited anterior
negativities; however, the response to the final verbs of scrambled wh-questions was statistically tenuous.

One dilemma that we face here is that scrambled demonstrative questions yielded more robust and reliable effects than did scrambled wh-questions. We have no ready explanation for why this should be the case. However, our acceptability ratings suggested to us that these differences were unlikely to have been caused by differences in acceptability. Scrambled demonstrative questions received an acceptability rating very similar to scrambled wh-questions, namely 2.8 versus 2.7, and there was no statistically reliable difference between the two. At this point, the reason for the quantitative differences in reliability of effects between scrambled demonstrative and wh-questions is not apparent, and we will continue to investigate it. However, since both scrambled conditions elicited similar (L)AN effects relative to the corresponding in-situ conditions, we shall treat these as equivalent in what follows.

Another concern is that LAN effects have typically been reported at post-gap positions in previous studies (Kluender & Kutas, 1993a,b; King & Kutas, 1995) rather than at pre-gap positions. In the present study, we obtained somewhat the opposite pattern of results. The head noun of the subject NP (‘adventurer-NOM’) certainly preceded the gap, and elicited both LAN and P600 effects. The adverb (‘finally’), which could have either preceded or followed a gap, also elicited both LAN and P600 effects. However, the final verb (‘discovered-Q’), which must have followed the gap, instead elicited bilateral anterior negativity that continued throughout the epoch. What should be made of this pattern of effects is unclear, and for lack of a better argument
we will merely assume that if P600 effects can apparently be elicited at both pre-gap (Kaan et al., 2000) and post-gap positions (Phillips et al., 2001; Phillips et al., 2003) across studies, there is no obvious reason why LAN effects should not exhibit the same variability.

This raises the further issue of the precise nature of the phasic (L)AN effects reported around the gap region. Were these merely shorter time slices of (an increase in) the longer-lasting slow negative potential ranging across several sentence constituents, or were they truly phasic effects specifically elicited by particular words in the string over and above the slow potential effect? The answer to this question will vary according to sentence position.

First, with regard to LAN effects in response to the subject head noun (‘adventurer-NOM’) and the adverb (‘finally’) of scrambled conditions, the answer must be that these were phasic effects: these anterior negativities were truly lateralized to the left hemisphere (Figures 2.6 and 2.7), as evidenced by interactions between condition, anteriority, and hemisphere in the temporal array that just missed significance. The slow anterior potential effect (Figure 2.4), on the other hand, was clearly symmetric, showing interactions only with anterior to posterior electrode position, but not with hemisphere. If the phasic effects at the subject head noun and the adverb were merely the result of (an increase in) the slow anterior negative potential, one would expect them to be bilateral as well.

Second, the negativity to the subject head noun (‘adventurer-NOM’) in scrambled demonstratives (Figure 2.6) showed every sign of being a phasic effect: at
left frontal electrodes (Fp1, F3, and F7), the ERPs to the subject head noun of the two conditions aligned almost perfectly at the N100 and P200 peaks, subsequently diverged in the N400 peak, and then realigned in the N100 and P200 components of the following word. If this difference were merely due to pre-existing differences in the slow negative potential, it would not show this realignment of componentry at the onset of the adverbial ‘finally’. Note that even though the second argument regarding componentry realignment does not hold for the LAN effects in response to the adverbial (Figures 2.6 and 2.7), the first argument regarding left-lateralization does, and we believe that is sufficient evidence for distinguishing this effect from the slow negative potential effects.

At the sentence-final verb, however, the left-lateralization argument also fails: the phasic effects of anterior negativity elicited in response to the final verb of scrambled sentences were not reliably left-lateralized. Thus the AN to the sentence-final verb could indeed have been caused by the slow negative potential seen in response to earlier sentence positions (Figure 2.4), and as such would not itself be a gap-filling effect. Under this scenario, we would be looking at something entirely different from gap-filling at the final verb position.

2.5.3.3 Post-gap AN

Before addressing what the AN to the final verb actually indexed, we can say with a fair degree of confidence what the effect at sentence end was not. It was not an anticipatory response to a following comprehension question. We say this for two reasons. First, the majority (80%) of stimulus sentences that participants saw did not
have comprehension questions associated with them, and there would be no reason to expect any difference between the scrambled and in-situ conditions along this dimension, as both were followed by the same total number of comprehension questions, i.e. 20% of the time. Moreover, the anterior negativity was unlikely to have been a mere sentence wrap-up effect as seen in previous ERP studies, which have been reported as an N400 difference (Osterhout & Holcomb, 1992; Hagoort, Brown, & Groothusen, 1993) or more recently as a widely-distributed sentence-end negativity (SEN) (Friederici & Frisch, 2000). Neither the morphology nor the distribution of our sentence-final effects was consistent with that of an N400 or SEN effect (see Figures 2.8 and 2.9). This ERP effect at sentence end therefore seems more likely to have something to do with the scrambling of prior constituents.

As discussed earlier, the bilateral slow anterior negative potentials elicited between filler and gap in scrambled conditions could simply have been an effect of holding an extra NP object in working memory at the point of comparison (O S __V vs. S O V), rather than a scrambling effect per se. However, the comparisons undertaken at the sentence-final verb position of the scrambled conditions provide clearer evidence that scrambling operations have their own associated processing costs independent of the number of NP arguments currently being held in the memory store. The sentence-final effects of anterior negativity in response to scrambling were not confounded by comparing different sentence positions or different lexical items, and cannot be attributed to different numbers of NP arguments being held in working memory across conditions either. By this point, the parser would have to have held
the same number of NPs in both conditions, and the negative anterior difference at sentence end can therefore only have been a consequence of scrambling in the present study.

In their self-paced reading time study, Mazuka et al. (2002) found that the subject of scrambled O S __V configurations took longer to read than the object of canonical S O V configurations, which also indicated that scrambling itself incurs a processing cost, independent of the number of NPs that must be held in working memory. Pechmann et al. (1994) and Rösler et al. (1998) likewise demonstrated costs of scrambling in acceptability, comprehension response time, and ERP studies using similar comparisons in German. However, here again we face a discrepancy with the results reported by Mazuka et al. (2001) in their ERP study of Japanese scrambling. As discussed earlier, Mazuka et al. (2001) elicited no slow anterior negative potential differences between filler and gap. They likewise elicited no phasic LAN effects at gap positions, and no differences of anterior negativity at the sentence-final verb of scrambled sentences. We surmise that the explanation for the lack of phasic LAN effects and of anterior negativities at sentence end will be the same as that for the lack of slow negative potentials in their data, namely, differences in the length of the stimulus materials. In other words, we believe it is consistent to claim that all effects of negative voltage deflection in our data were tied in one way or another to the verbal working memory demands of our scrambled sentences: the slow negative potentials to the entry and maintenance of a filler in working memory, the phasic LAN effects to
the retrieval of this filler from working memory for purposes of gap assignment, and
the anterior negativities at sentence end to continuing working memory demands.

2.5.3.4 LAN vs. P600

We think that this interpretation of our results can also help to make sense of a
long-standing conundrum in the ERP literature regarding the functional significance of
LAN vs. P600 effects. Recall that in our data, both effects were elicited in response to
the subject head noun (‘adventurer-NOM’) of our scrambled demonstrative condition.
We have interpreted these in terms of the costs associated with retrieving a filler from
verbal working memory for purposes of gap-filling (LAN) and syntactically
integrating this filler-gap relationship in the ongoing parse (P600). This is far from
the first time that this particular pattern of results has been observed, however. In fact,
the very first study of P600 effects (Osterhout & Holcomb, 1992) reported the exact
same combination of effects (i.e. LAN plus P600) in response to *was in the sentence
*The broker hoped to sell the stock was sent to jail. As originally suggested in
Kluender and Kutas (1993b: 622-624), the P600 can in this case be interpreted as an
index of syntactic processing difficulty, while the LAN can be interpreted as a
desperate (and ultimately futile) attempt on the part of parser to seek an appropriate,
previously occurring discourse referent in verbal working memory to assign as the
subject of the verb was. This interpretation is also consistent with the results reported
in Urbach (1993) in response to was in the garden path sentence The cook helped in
the kitchen was busy. In this instance, the elicited LAN effect would index a (this
time) successful attempt on the part of the parser to locate an appropriate, previously
occurring discourse referent for the verb was in verbal working memory, while the P600 effect would reflect the syntactic processing difficulty involved in reanalysis of the string helped in the kitchen as a reduced relative clause. If this account is accurate, it would provide a fortuitous index of simultaneous but dissociable working memory and syntactic integration costs in the ERP record (cf. Gibson, 1998, 2000).

2.5.4 Larger Implications

The pattern of results obtained in the present study is incompatible with a model of sentence processing based on head-driven parsing (e.g., Pritchett, 1992). It is possible to say that the slow anterior negative potentials between filler and gap index the presence of an extra NP filler before thematic role assignment and attachment have taken place at the final head, and that the sentence-final AN effects index the extra effort on the part of the parser in attaching scrambled NPs, possibly because recalling the object and the subject in reverse order for thematic role assignment incurs an extra processing cost. However, phasic LAN and P600 effects elicited at the subject head noun and adverb positions are difficult to account for under this model, as it predicts no effects of gap-filling before the sentence-final verb. In addition, the pattern of ERP effects in this study cannot be attributed to thematic ambiguity of the filler, as suggested by King and Kutas (1995) in their study of relative clauses in English. Since Japanese is a case-marking language, the grammatical functions of the two argument NPs, and consequently their respective thematic roles, were easy to determine. Although Japanese exhibits cases of unusual
linking between a case marker and a grammatical function (e.g., dative subject, nominative object), we never presented examples of such sentences.

What type of model can then account for our results? As discussed above, the pattern of results would support a model of incremental (e.g., Inoue & Fodor, 1995) rather than head-driven (e.g., Pritchett, 1992) processing. But in addition to that, what other processing model would be most compatible with the effects at specific sentence positions in the present study? At a minimum, the pattern of results obtained can be said to have something to do with the non-canonical order of scrambling. “Canonical word order” may be defined as the most statistically frequent and contextually neutral word order (Comrie, 1981). In Japanese, scrambled sentences are reported to occur very infrequently, less than 1% of the time in transcriptions of magazine articles in various registers of formality (N = 2,635: Yamashita, 2002). In regard to contextual neutrality, as discussed earlier, Yamashita reported that 95% of the scrambled sentences in her corpus could be accounted for either by reference to the immediately preceding context or by “heaviness” in terms of syntactic complexity. Thus it seems safe to conclude that scrambled word orders are indeed non-canonical in Japanese, and that this fact played a role in the ERP results elicited in this study.

Of course, there are several models that can explain the processing of non-canonically ordered sentences, and these can be categorized according to how predominant a role they assign to syntactic structure during processing. Several accounts link processing costs more or less directly to the hierarchical phrase structural properties of stimulus sentences. Such accounts would include the active
filler strategy, in which the parser tries to dispose of a filler as soon as possible (Frazier & Clifton, 1989), the minimal chain principle, in which the parser tries to minimize the length of a syntactic chain (De Vincenzi, 1991), and a more generalized memory strategy, in which the amount of syntactic structure kept in memory influences parsing times (Schlesewsky, Fanselow, Kliegl, & Krems, 2000). Other accounts link processing costs to the frequency of occurrence of particular syntactic constructions. For instance, Mitchell, Cuetos, Corley, and Brysbaert (1995) proposed the tuning hypothesis to account for a difference in relative clause attachment preference in English and Spanish. They argued that parsing preferences reflect statistical biases in the linguistic input of subjects, i.e. that structural parsing decisions are influenced (or tuned) by subjects’ prior contact with the relative frequencies of various syntactic structures as reflected in corpus studies.

However, problems have been pointed out in relation to models based on statistical frequency. It is not always the case that frequency can be correlated with comprehenders’ preferences: there have been several studies suggesting that frequency in corpora do not always match behavioral preferences. Most generally, various problems have been pointed out regarding the way in which corpus studies have been carried out (Roland, Elman, & Ferreira, 2003). This empirical problem can then easily lead to behavioral mismatches with claimed corpus frequencies. Gibson, Schütze, and Salomon (1996) investigated preferences in NP attachment in English and showed that while both the Brown and Wall Street Journal corpora indicated an attachment preference in the order high < middle < low, their acceptability rating
experiment indicated a preference in the order middle < high < low. Schlesewsky et al. (2000) reported that whereas their behavioral (sentence completion and reading time) studies indicated a clear preference for a subject over an object in locally ambiguous German wh-questions, their corpus data showed roughly equivalent frequency counts between subject and object wh-questions. In addition, Miyamoto and Takahashi (2002a) showed that the preference for shorter filler-gap distances shown by their subjects in a reading time experiment on Japanese scrambled sentences was not directly reflected in corpus frequencies.

Aside from the empirical problem of how corpus frequencies are established, there is the additional interpretational problem that the comprehension system might be controlled by some mechanism distinct from what drives the frequencies in a production corpus. As Mazuka et al. (2002) point out, frequency of occurrence alone cannot account for increased processing costs at a particular sentence position without appealing to some notion of sentence structure. It may be the case that once the parser recognizes a non-canonical word order (at the scrambled object NP position or somewhere at the subject NP position in the present experiment; see Section 2.5.1), it continues to bear the resulting processing costs all the way through to sentence end, and that the slow anterior negative potentials between filler and gap, and perhaps the AN effect at the final verb position as well, index such an effect. However, the P600 effects following the filler and at the gap for scrambled demonstrative questions, and the phasic LAN effects at the gap position for scrambled demonstrative and wh-questions require additional reference to sentence structure position.
Thus any model that would account for our data would have to be able to refer to some notion of sentence position and the syntactic integration of a filler-gap dependency. In this sense, both linear and hierarchical models of sentence structure would suffice. In a linear structural model, the parser would simply be assumed to be attempting to put the scrambled NP object back into its canonical SOV word order position. The motivation for this mental operation might be that the human parser prefers canonical word order as the path of least resistance. An alternative option is a processing model based on hierarchical phrase structural representations. Such a model would argue that the parser is an autonomous system whose primary goal is to construct phrase structural representations for incoming words. This concept was proposed early on in the history of psycholinguistics (e.g. Kimball, 1973; Fodor, Bever, and Garrett, 1974; Frazier & Fodor, 1978), and the structural processing models mentioned earlier (the active filler strategy, the minimal chain principle, and the generalized memory for syntactic structure strategy) are based on this concept.

With the present data, we cannot conclusively determine whether a linear structure model is better than a hierarchical structure model or vice versa, although there would clearly be some advantages of economy to adopting a hierarchical model. A hierarchical model can directly justify the preferences in structurally based ambiguity (e.g., the minimal attachment principle: Frazier & Fodor, 1978) or a processing cost based on a structural distance (O’Grady, 1997), even though this is admittedly not an advantage directly related to the processing of filler-gap dependencies.
To illustrate how scrambled sentences would be processed based on a hierarchical phrase structural representation that also relies on syntactic movement, the phrase structures for the scrambled and in-situ demonstrative conditions used in the present study following the movement analysis by Saito (1985)\textsuperscript{11} are given in Figure 2.10.

\begin{figure}[h]
\centering
\begin{tikzpicture}
  \node {CP}
  child {node {IP}
    child {node {NP_{i} that-ACC}
      child {node {NP}
        child {node {I’ the reckless adventurer-N}
          child {node {AdvP finally t_i V discover}}
        }
        child {node {VP PAST}}
      }
    }
    child {node {C Q}}
  }
  child {node {Q}}

  \node {CP}
  child {node {IP}
    child {node {NP}
      child {node {I’ the reckless adventurer-N}
        child {node {AdvP finally}}
        child {node {VP PAST}}
      }
    }
    child {node {C Q}}
  }

\end{tikzpicture}
\caption{Phrase structure trees of stimulus sentences under the movement analysis of scrambling by Saito (1985).}
\end{figure}

Under this analysis, forming a chain between ‘that-ACC’ and the gap (indicated by the trace t) would require keeping the head of the chain in working memory (Fiebach et al., 2001) until the tail of the chain is reached, eliciting anterior negative slow potentials. The P600 effects following the filler (at ‘the’ in Figure 2.10.a) and

\textsuperscript{11} These phrase structures use an updated tree convention different from that used in Figures 2.1 and 2.2 in order to accommodate the question particle, which will be discussed in detail in the following
preceding the gap (at ‘adventurer-N’ and ‘finally’ in Figure 2.10.a) index the establishment and integration of a syntactic chain. The phasic LAN effects in the pre-gap region (at ‘adventurer-N’ and ‘finally’ in Figure 2.10.a) are related to the retrieval of the filler, or the activation of the trace-antecedent relationship, from working memory.

Relying on a hierarchical phrase structural model of scrambling also has the advantage that it accounts for other syntactic phenomena that have no explanation under a linear structural model. For example, both the movement analysis of scrambling shown in Figure 2.10, as well as the non-configurational, “flat” structural analysis of Hale (1980) and Farmer (1980) shown in Figure 2.1, are compatible with the basic problem of non-canonical word order under scrambling of objects. Under a non-configurational, “flat” structural analysis, there is no trace-antecedent relationship created via movement of the object to a higher phrase structural position. Nonetheless, one can imagine ways in which a non-configurational analysis could still accommodate the ERP results of the current study, for example by attempting to recreate the flat structure shown in Figure 2.1.a from the flat structure in Figure 2.1.b, or perhaps by postulating mental comparisons across simultaneous parallel flat structures (Figure 2.1.a and 2.1.b, cf. Goodall 1984). However, while such mechanisms might be invoked to account for the present data, we know from Saito’s (1985) early work that the flat structures in Figure 2.1 cannot account for facts of Japanese anaphoric binding indicating that the scrambled object must be in a higher position in a hierarchical phrase structural representation, and most specifically
outside the verb phrase. Since these independent facts of binding clearly point to the existence of a verb phrase in Japanese, reasons of economy would dictate the adoption of a hierarchical phrase structural representation underlying the ERP indices of scrambling in the present study as well – while remaining completely agnostic on the issue of whether the hierarchical antecedent-trace relationship between a scrambled object and its gap were formed by movement or not.

Thus both linear and hierarchical structural approaches can account for the slow potentials related to holding a filler in working memory, for P600s in response to the syntactic integration of a filler-gap dependency, and for phasic LANs in response to the retrieval of the filler from working memory for purposes of gap assignment. Independent reasons suggest however that it is more economical to conceptualize the relationship between a scrambled object and its trace or gap position as a hierarchical one.

At the same time, a model that ties sentence processing effects directly to hierarchical phrase structural representations in the grammar may have a harder time accounting for the sentence-final AN effects of this study. In this case, a statistically based frequency account of structural processing could account for the sentence-final AN effects as a continuation of the slow anterior negative potentials between filler and gap by claiming that non-canonical word order causes a continuous processing load all the way through to sentence end. Models of sentence processing that refer directly to particular structural positions in a hierarchical phrase structure tree (e.g. traces of movement, the tails of syntactic chains, phrase structural projections, etc.) do not
directly predict such end-of-sentence effects, as at that point in the sentence, gap-filling has already occurred. So under such an account it is not obvious why ERP effects should persist through to sentence end in both head-initial and head-final languages.

It has been known since early ERP studies of filler-gap dependencies (Kluender & Kutas, 1993a; King & Kutas, 1995) that while ERP signals indicating extra processing costs generally begin after the filler has been entered into the parser, they continue well past the gap position as residual effects. Recall however from Section 2.5.1 that Mazuka et al. (2001) reported no anterior negativity effects (in either phasic or slow potential forms) between filler and gap, or anywhere else, including the sentence-final verb position in their scrambled sentences, and that we concluded that this was because their stimulus sentences were too short to elicit any visible working memory effects. This is in contrast with the P600 effects that were elicited at the object and subject NP positions of the scrambled condition, which were consistent with our data despite differences in sentence length. This seems to suggest that the slow anterior negative potentials (both between filler and gap and at the sentence final-verb position) elicited in the present study were not directly related to phrase structural processing of scrambled word order; if anterior negativity effects directly indexed the phrase structural processing of non-canonical word order, they should be elicited even for a short sentence, as with the P600 effects. The fact that Mazuka et al.’s short scrambled sentences did not elicit sentence-final AN seems to suggest that the sentence-final AN effects found in the present study were due to
subjects having had to hold a filler for a long time. That is, the sentence-final AN was an index of residual working memory processes, rather than an index of the structural challenges posed by processing non-canonical order per se.\textsuperscript{12}

To summarize, it would be hard to explain the entire pattern of results without using some notion of structural position, which can be based on either a linear or a hierarchical model of Japanese sentence structure. Both linear and hierarchical models can account for the formation of a syntactic dependency between filler and gap, yet for independent reasons of syntactic analysis, a hierarchical model seems clearly preferred. Models that tie processing effects directly to particular phrase structural positions in a hierarchical syntactic representation do not directly predict that anterior negative potentials should persist to sentence end, while models based on syntactic frequency preferences seem to offer a more straightforward account of these ERP effects, particularly in view of the very low frequency of occurrence of scrambling in Japanese corpus studies. However, if these slow anterior negative potentials are taken as an index of either generalized or syntactic working memory, rather than of phrase structural processing per se, models that rely on hierarchical phrase structural representations also seem perfectly compatible with the end-of-sentence effects.

\textsuperscript{12} This interpretation of linking anterior negativity to general working memory load goes along with several reports of slow anterior negative potential effects that are not linked to any syntactic structural difference with both sentential (Münte, Schiltz, & Kutas, 1998) and non-sentential (Ruchkin, Johnson, Canoue, & Ritter, 1990; Ruchkin, Johnson, Grafman, Canoue, & Ritter, 1992) stimuli (see Chapter 4 for details).
2.6 Conclusion

Despite substantial differences in language family, orthography, basic word order, and the potential for syntactic ambiguity, this study uncovered no entirely new or unexpected ERP components in response to Japanese language stimuli. One may thus conclude that language-related ERP components are universal across languages: despite some discrepancies, the results of this experiment in general followed similar patterns seen previously in English and German, and now also in Japanese. The word immediately following scrambled demonstratives elicited a P600; this was similar to Mazuka et al.’s (2001) report of a P600 at the scrambled NP direct object of \( OS \_V \) sentences. Slow anterior negative potentials were elicited by the words intervening between scrambled elements and their canonical positions, similar to the slow potential effects reported in King and Kutas (1995), Münte et al. (1997), Kluender and Münte (1998), and Fiebach et al. (2001). At the head noun of the subject NP, both phasic LAN and P600 effects were seen. The latter were reminiscent of P600 effects at the pre-gap verb position of wh-movement sentences (Kaan et al., 2000). Phasic LAN effects have not previously been reported at pre-gap positions, but may well be related to the post-gap LAN effects reported in English (Kluender & Kutas, 1993a, b; King & Kutas, 1995). In addition, an effect of anterior negativity persisted at the sentence-final verb, as has also been the case in earlier English studies. The pattern of results was incompatible with a heard-driven parser model (Pritchett, 1992), and suggested a incremental processing model (e.g., full attachment model, Inoue &
Fodor, 1995). Furthermore, the results were compatible with several processing models pertaining to filler-gap dependencies.
Chapter 3: ERP Measures of Mono-Clausal Japanese Wh-Questions

3.1 Introduction

3.1.1 Japanese: A Wh-in-Situ Language

This chapter examines the processing of mono-clausal Japanese wh-questions. Consider the English sentences in (3.1). In (3.1b), the wh-object what is displaced to the beginning of the clause compared to the non-wh object pizza in (3.1a).

(3.1) a. Did Calvin bring pizza?

b. What did Calvin bring __?
        FILLER             GAP
        |___________________|

Recall that in Chapter 2 we reviewed several ERP studies on the processing of filler-gap dependencies (e.g., Kluender & Kutas, 1993a,b; Kluender & Münte, 1998; Fiebach, Friederici, Schlesewsky, 2001). These ERP studies argue that associating the displaced wh-filler (what in (3.1b)) with its gap position (indicated by underlining in (3.1b)) increases verbal working memory load, and that this processing cost is reflected in an ERP component known as left anterior negativity (LAN). These studies have been done in English and German, so-called “wh-movement” languages that require wh-elements to be displaced to the beginning of a clause (except in the case of an echo question like Calvin brought what?).

Unlike English and German, Japanese is a “wh-in-situ” language where wh-words stay in the same canonical SOV (subject-object-verb) position as their non-wh counterparts. As shown in (3.2), ‘pizza’ and ‘what’ stay in the same position.
(3.2)  

a. カルビンが ピザを 持ってきたんですか。
   Calvin-ga pizza-o mottekita-ndesu-ka.
   Calvin-NOM(INATIVE) pizza-ACC(USATIVE) brought-POLITE-Q(QUESTION)

‘Did Calvin bring pizza?’

b. カルビンが 何を 持ってきたんですか。
   Calvin-ga nani-o mottekita-ndesu-ka.
   Calvin-NOM what-ACC brought-POLITE-Q

‘What did Calvin bring?’

Although it is also possible in Japanese to displace objects to the beginning of the clause, this involves another process called scrambling as discussed in Chapter 2. Scrambling can be applied to both wh- and non-wh-elements, as shown in (3.3).

(3.3) Scrambled Word Order

a. ピザを カルビンが 持ってきたんですか。
   pizza-o Calvin-ga mottekita-ndesu-ka?
   pizza-ACC Calvin-NOM brought-POLITE-Q

‘Did Calvin bring pizza?’

b. 何を カルビンが 持ってきたんですか。
   nani-o Calvin-ga mottekita-ndesu-ka?
   what-ACC Calvin-NOM brought-POLITE-Q

‘What did Calvin bring?’

While displacement is optional, Japanese wh-words always require a question [Q] particle *ka* or *no* (meaning ‘whether’) at the end of the clause, as in (3.4a).\(^1\) Wh-questions without a Q-particle are ungrammatical in Japanese, as shown in (3.4b).

---

\(^1\) The exception is when there is a rising intonation in speech, which can be interpreted as a prosodic version of a Q-particle.
(3.4)

   Calvin-NOM  what-ACC  brought-POLITE-Q

‘What did Calvin bring?’

b. *Calvin-ga  nani-o  mottekita-ndesu-yo.
   Calvin-NOM  what-ACC  brought-POLITE-you.know

‘*Calvin brought what.’

In addition, this Q-particle *ka* determines the interrogative scope of a wh-element.

Interrogative scope can be defined as the domain of the sentence that is being questioned.

The surface position of moved wh-elements in wh-movement languages, such as English and German, transparently indicates their interrogative clausal scope relations in the sentence. For instance, both (3.5a) and (3.5b) consist of two clauses, matrix and embedded. The wh-element *what* can be placed either at the beginning of the embedded clause as in (3.5a), yielding an embedded clause wh-question (traditionally termed an “indirect question”)\(^2\), or at the beginning of the matrix clause as in (3.5b), yielding a matrix clause wh-question (also termed a “direct question”).

Thus, the logical answer to an embedded clause wh-question like (3.5a) would be *yes* or *no*, since no element in the matrix clause is questioned by a wh-word. On the other hand, the logical answer to a matrix clause wh-question like (3.5b) would be the referent of the wh-word (*pizza* in this case).

---

\(^2\) More strictly speaking, (3.5a) is both an embedded clause wh-question and a (matrix clause) yes/no-question. The embedded clause wh-question only would be *Hobbes said what Calvin brought*. However, the form shown in (3.5a) is used because it is more parallel to the ERP stimuli discussed later.
(3.5) Wh-Movement Languages (e.g., English)

a. Embedded clause wh-question

[Did Hobbes say [what Calvin brought __ ]]? Logical Answer: Yes, he did.

b. Matrix clause wh-question

[What did Hobbes say [Calvin brought __ ]]? Logical Answer: Pizza.

In Japanese, instead of the surface position of a wh-element, the position of its related Q-particle indicates its interrogative clausal scope. As shown in (3.6), while the wh-elements in the embedded clause wh-question (3.6a) and the matrix clause wh-questions (3.6b) stay in the embedded clause, the Q-particle *ka* appears at the end of the interrogative clause, either embedded (3.6a) or matrix (3.6b).3

---

3 Although this is a standard analysis in the Japanese wh-literature (e.g., Nishigauchi, 1990; Watanabe, 1992), it has been claimed that sentences like (3.6a) also can have a matrix wh-reading, as in ‘What did Hobbes say whether Calvin brought?’ (Takahashi, 1993; Hirotani, 2003). However, this reading seems to be strongly dispreferred (I myself cannot get the reading) and requires a specific prosodic cue (lack of a pitch reset; see Chapter 4 Section 4.5.1.2). Hirotani (2003) shows that without such a prosodic cue, there was a strong preference (84%) for an embedded clause wh-question response, as in ‘Yes, he did’. Even with the prosodic cue, about half (43%) of her wh-questions like (3.6a) elicited an embedded clause response. Our sample production data show that the embedded clause wh-questions in our stimuli are unlikely to have such a prosodic cue, indicating the embedded clause wh-question reading as intended (see Chapter 4 Section 4 for more details). In addition, as noted in footnote 1, it is not clear how much of the matrix wh-question response (as in ‘Pizza.’) in Hirotani’s (2003) data was pragmatics-driven. Here we proceed with our discussion assuming that an embedded Q-particle indicates an embedded clause wh-reading.
(3.6)
a. Embedded clause wh-question

\[ \text{ホッブスはカルビンが何を持ってきたか言ったんですか。} \]

\[ \text{Hobbes-wa Calvin-ga nani-o mottekita-ka itta-ndesu-ka]? } \]

‘Did Hobbes say what Calvin brought?’

b. Matrix clause wh-question

\[ \text{ホッブスはカルビンが何を持ってきたと言ったんですか。} \]

\[ \text{Hobbes-wa Calvin-ga nani-o mottekita-to itta-ndesu-ka]? } \]

‘What did Hobbes say Calvin brought?’

For matrix clause wh-questions like (3.6b), a non-question particle to ‘that’ is attached to the embedded verb.

Thus as shown in (3.7), a wh-element is displaced to the left of its gap and marks its scope in English, while in Japanese a Q-particle is placed to the right of the corresponding wh-element and marks its scope.

(3.7)
a. English

\[ \text{What did Hobbes say [Calvin brought __]?} \]

b. Japanese

\[ \text{Hobbes-wa Calvin-ga nani-o mottekita-to itta-ndesu-ka]? } \]

‘What did Hobbes say Calvin brought?’
In this sense, the wh-scope marking system in Japanese may look like a mirror image of the wh-scope marking system in English.

3.1.2 Theoretical Claims about Wh-in-Situ

Given the above syntactic differences, theoretical claims have been made regarding wh-in-situ that can be roughly classified into two categories: LF-movement and Q-coindexation. Both claims are based on the assumption that a wh-element is a variable that needs its interrogative scope to be licensed. In wh-movement languages, a wh-element moves to its scope position and binds its trace, which is assumed to be at the gap position as shown in (3.8) (binding represented by coindexation). The interrogative scope of a wh-element is licensed in this way.

(3.8)

a. What did Calvin bring t(race)?

b. [Did Hobbes say [what Calvin brought t]]?

c. [What did Hobbes say [Calvin brought t]]?

The LF-movement analyses argue that although wh-movement and wh-in-situ languages yield different surface word orders, they occur in the same order at some level of linguistic representation. The Government and Binding Theory and the Minimalist Program argue for a level of representation called Logical Form (LF). LF is said to be an intermediate level between syntax and semantics, where quantificational elements, such as wh-elements, occupy their scope positions (Chomsky 1976; May 1977, 1985; Aoun, Hornstein & Sportiche 1981; Jaeggli 1982).
Some linguists also argue that at LF, wh-elements in wh-in-situ languages are displaced to their clausal scope positions, just as in wh-movement languages. This process is called LF-movement (Huang 1982; Lasnik & Saito 1984; Nishigauchi 1990). Thus at LF, wh-in-situ sentences like those in (6) would have the representations shown in (3.9)⁴.

(3.9) LF Representation

a. Embedded clause wh-question

何をホッブスはカルビンが持ってきたか言ったんですか。

[Hobbes-wa [nani-o Calvin-ga mottekita-ka] itta-ndesu-ka]?

'[Did Hobbes say [what Calvin brought __ ]?]

b. Matrix clause wh-question

何をホッブスはカルビンが持ってきたと言ったんですか。

[nani-o Hobbes-wa [Calvin-ga mottekita-to] itta-ndesu-ka]?

'[What did Hobbes say [Calvin brought __ ]?]

⁴ Huang (1982) proposes that wh-elements in wh-in-situ languages like Chinese undergo leftward movement at LF, as shown in (ib).

(i) a. Surface structure

Zhangsan yiwei Lisi mai-le shenme?
Zhangsan thinks Lisi bought what

'What does Zhangsan think Lisi bought?'

b. LF representation

[shenme Zhangsan yiwei [Lisi mai-le __ ]] what Zhangsan thinks Lisi bought

On the other hand, Lasnik and Saito (1984) and Nishigauchi (1990) actually argue that wh-elements are displaced to the right, next to the question particle. However, here we assume leftward movement simply for expository purposes.
Notice the order of wh-words in the Japanese sentences: the wh-word *nani* ‘what’ in the embedded question (3.9a) is at the beginning of the embedded clause, while the wh-word *nani* ‘what’ in the matrix clause question (3.9b) is at the beginning of the matrix clause. This reflects ‘what’ having been raised to its interrogative scope position at LF, and in effect, yielding an embedded clause wh-question in (3.9a) and a matrix clause wh-question in (3.9b).

The Q-coindexing analyses focus on the relationship between wh and Q. Nishigauchi (1990) and Cheng (1991) argue that some wh-in-situ languages require Q-particles that determine the quantificational (including interrogative, existential and universal) force of wh-elements. Further, several linguists (Baker, 1970; Pesetsky, 1987; Watanabe, 1992; Aoun & Li, 1993; Cole & Hermon, 1994, 1995) propose the mechanism of binding by a Q-operator. In this mechanism, the scope of a wh-element is represented by a coindexing Q-operator rather than a moved wh-element. As shown in Figure 3.1, a covert Q-operator is postulated in the spec of CP, with an overt/covert Q-particle in C linked by spec-head agreement. Then the operator, either base-generated (Baker, 1970; Pesetsky, 1987; Cole & Hermon, 1994, 1995) or moved (Watanabe, 1992; Aoun & Li, 1993), is said to bind the wh-element.

---

5 Cheng (1991) proposes that languages with ambiguous wh-words, i.e., wh-words with interrogative, existential and universal interpretations, (e.g., Japanese, Korean) require the presence of an overt Q-particles, while languages with non-ambiguous wh-words (e.g., Hindi, Bahasa Indonesia) allow covert Q-particles.
These two types of analyses appear different at first glance, yet they share a common property: the scope of a wh-in-situ needs to be licensed by something in the spec of its interrogative clausal CP. This “something” can be either the moved wh-element itself or a moved/base-generated operator that is linked with its Q-particle.\(^6\)\(^7\)\(^8\)

Thus the LF-movement and Q-coindexation analyses can be subsumed into a licensing dependency between a wh-in-situ and something in the CP that marks its scope (or COMP, to use a term that covers both Spec and C).

### 3.1.3 Processing of Japanese Wh

There is experimental evidence suggesting some type of relationship between Japanese wh-elements and Q-particles. In their ERP study, Nakagome et al. (2001)

---

\(^6\)Nishigauchi (1990) actually has both LF-movement and Q-coindexation approaches in his proposal, suggesting the compatibility of the two types of analysis. He assumes LF-movement of wh-in-situ and argues that a moved wh-in-situ is governed and bound by the Q-particle at LF and gets its quantificational force licensed. Thus in this case, a wh-in-situ moves once to COMP, and then gets its scope licensed there. However, in a larger view, this can still be considered as a licensing dependency between a wh-in-situ and its COMP.

\(^7\) Following the restrictions under the Minimalist program (Chomsky, 1995) that no formal features are strong, Hornstein (1995) argues against wh-movement at LF and suggests that wh-in-situ might be “functionally interpreted” with a null pronoun that is coindexed with it.

\(^8\)Pesetsky (2000) argues that all wh-phrases require overt movement and that the in-situ versus fronted contrast is a question of whether to pronounce the foot or the head of the chain. This analysis can also be subsumed under the generalization that something in the spec of CP binds wh-in-situ.
examined the processing of mono-clausal wh-questions with and without Q-particles, as shown in (3.10), at the sentence-final verb position.

(3.10)

a. 動物園で 何を 見たか。
   Dobutsuen-de nani-o MI-TA-KA.
   zoo-LOC what-ACC see-PAST-[+Q]

   ‘What did (you) see in the zoo?’

b. 動物園で 何を 見たよ。
   *Dobutsuen-de nani-o MI-TA-YO.
   zoo-LOC what-ACC see-PAST-[+Q]

   *(I) saw what at in zoo’

The results showed a P600 (more positivity between 400 and 700 ms post-stimulus onset predominantly at bilateral posterior sites) in response to wh-Q violations (8b). Nakagome et al. (2001) attributed the P600 to syntactic processes related to the operation of movement.9

In a self-paced reading experiment, Miyamoto and Takahashi (2001, 2002b) compared bi-clausal Japanese matrix and embedded clause wh-questions, as well as their structurally equivalent non-wh-counterparts, at the embedded verb position, as shown in (3.11).

---

9In this particular case, following the Minimalist Program (Chomsky 1995; 1998), they suggested that wh-features on wh-words have to be moved to the clausal scope position and be checked by a [+Q] particle at LF, and that the inability of the anomalous sentences with a [-Q] particle to do so resulted in the P600 effect. Also see Nakagome, Takazawa, Kanno, Hagiwara, Nakajima, and Ito (1999) for preliminary data and more discussion on possible implications for the Minimalist Program (Chomsky 1995; 1998).
They found that at the embedded verb position, matrix clause wh-questions (3.11b) were read more slowly than embedded clause wh-questions (3.11a). They argued that since Japanese wh-words, unlike those in English, are not used to form relative clauses, a wh-word in a sentence signals a Q-particle to appear\(^\text{10}\). Miyamoto and Takahashi argued that a wh-word creates an expectation for interrogative typing of its clause by a Q[uestion]-particle (Cheng, 1991) while a non-wh-word creates an expectation for declarative typing of its clause by a non-Q-particle as the default. When this expectation is violated, it causes a longer reading time. Matrix clause wh-questions

\footnote{Strictly speaking, Miyamoto and Takahashi (2001: footnote 3) cite Nishigauchi (1990), who argues that a wh-phrase requires a particle that can determine its quantificational force.}
(3.11b) were read more slowly because they violated this expectation. They called this delay in reading times the Typing Mismatch Effect (TME).

Similarly, Aoshima, Phillips, and Weinberg (in press) observed a TME slowdown even when a wh-element has to be interpreted as scrambled out of the embedded clause, as shown in (3.12).

(3.12)

a. Scrambled, Question Particle (Embedded clause wh-question)

EMBEDDED VERB

どの生徒に 担任は 校長が 本を 読んだか 図書室で 司書に 言いました。
Dono-seito-ni tannin-wa [kocho-ga hon-o yonda】 toshoshitsu-de shisho-ni iimashita.
which-student-D teacher-T principal-N book-A read-Q library-at librarian-D said

‘The teacher told (to) the librarian at the library to which student the principal read a book’

b. Scrambled, Declarative Complementizer (Matrix clause wh-question)

EMBEDDED VERB

どの生徒に 担任は 校長が 本を 読んだと 図書室で 司書に 言いました。
Dono-seito-ni tannin-wa [kocho-ga hon-o yonda-to] toshoshitsu-de shisho-ni iimashita-ka.
which-student-D teacher-T principal-N book-A read-th at library-at librarian-D said-Q

‘To which student did the teacher tell (to) the librarian at the library that the principal read a book?’

They reported that the reading time at the embedded verb position was longer for matrix clause wh-questions (3.12b) than for embedded clause wh-questions (3.12a). The wh-element dono-seito-ni ‘which-student-D’ can be interpreted as a part of the embedded or matrix clause, as neither the embedded verb yonda ‘read’ nor the matrix verb iimashita ‘said’ requires a dative NP as its mandatory argument. However, the
result suggests that the parser prefers to interpret the scrambled wh-element as a part of the embedded clause and associate the wh-element with the embedded verb-Q position, although that would involve a clause external scrambling interpretation of the wh-element.11

The above studies suggest that there is some relationship between Japanese wh-elements and Q-particles. Although their interpretations differ (Miyamoto and Takahashi (2001, 2002b) and Aoshima et al. (in press) attributed their results to the parser’s expectation for a Q-particle, and Nakagome et al. (2001) to the syntactic operation of movement), the focus of these studies was on a Q-particle. This goes along with the theoretical claims discussed in Section 3.1.2 that wh-in-situ has to have its scope licensed by something in the corresponding interrogative COMP (either moved wh or Q-particle).

3.1.4 Predictions

The present experiment examined the processing of mono-clausal Japanese wh-questions. The mono-clausal design was chosen in order to investigate, as a first step, any effects of Japanese wh-processing in a simple structure. The aim of the present experiment was to examine the extent to which the neural processing of wh-questions in a wh-in-situ language, such as Japanese, shows similarities to the processing of wh-questions in wh-movement languages, such as English and German. More specifically, based on the above theoretical claims and experimental studies that suggest some type of relationship between a wh-element and its corresponding Q-

---

11 Aoshima et al. (in press) also had an in-situ version of this comparison, which also showed a TME at the embedded verb position.
particle, we hypothesized that, similar to filler-gap dependencies in English or German, there might be a dependency between Japanese wh-elements and their Q-particles. Further, we predicted that if there really is a dependency between a Japanese wh-element and its Q-particle *ka*, and if the wh-Q dependency is similar to a filler-gap dependency, we may see the same type of ERP components, such as LAN and slow negative anterior potentials (Kluender & Kutas, 1993ab; King & Kutas, 1995; Kluender & Münte, 1998; Fiebach et al., 2001), between Japanese wh-words and related Q-particles, as shown in (3.13).

(3.13) Wh-Q Dependency Hypothesis

a. Filler-gap dependency for wh-movement languages (e.g., English, German)

```
What did Calvin bring ___?
Filler GAP

(L)AN
```

b. Wh-Q dependency for wh-in-situ languages? (e.g., Japanese)

```
カルビンが 何を 持ってきたんですか。
Calvin-ga nani-o mottekita-ndesu-ka
Calvin-NOM what-ACC brought-POLITE-Q
WH Q

(L)AN?
```

‘What did Calvin bring?’

In addition, as a wh-in-situ needs its interrogative scope licensed by some mechanism, we hypothesized that there might be an interpretive process involved in calculating the scope of wh-in-situ. We further hypothesized that this “scope-calculation effect” might be seen at the Q-particle position that determines the scope
of a wh-in-situ. Thus, for mono-clausal wh-questions in the present experiment, such an effect was predicted at the sentence-final verb-Q complex as shown in (3.14).

(3.14) Wh-Scope Calculation Hypothesis

<table>
<thead>
<tr>
<th>Calvin-ga</th>
<th>nani-o</th>
<th>mottekita-ndesu-ka</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calvin-NOM</td>
<td>what-ACC</td>
<td>brought-POLITE-Q</td>
</tr>
</tbody>
</table>

'What did Calvin bring?'

However, as there is/was no known prior ERP study of grammatical wh-in-situ processing, it was hard to predict the ERP component(s) that might index such a process. It might well be LAN if the parser has to recall the words contained in the interrogative clause for scope calculation, which may tax working memory load. It might also be a P600 if we assume, along with Nakagome, et al. (2001), that the syntactic operation of movement, in this case wh-raising, involves such a component (see also Kaan et al. (2000) and Phillips, et al. (2001, 2003) for discussion of the P600 as the index of syntactic integration effect).

Further, we wanted to see whether there is any discernible relationship between particular brain responses and linguistic claims made with regard to wh-in-situ languages such as Japanese. Particularly, we wanted to see whether (and if so, how) the theoretical mechanisms proposed for licensing the scope of a wh-element would be reflected in brain responses.
3.2 Norming

A paper-and-pencil rating study on a subset of the experimental items was conducted to norm the experimental materials. Stimuli consisted of four conditions of mono-clausal questions, namely, (a) object wh-in-situ questions (hereafter called “wh-in-situ questions”), (b) yes/no-questions with demonstrative objects in-situ (hereafter called “yes/no-in-situ questions”\(^{12}\)), (c) scrambled object wh-questions (hereafter called “scrambled wh-questions”), and (d) yes/no-questions with scrambled demonstrative objects (hereafter called “scrambled yes/no-questions”\(^{13}\)), as shown in Table 3.1 (see also Appendix 1).

\(^{12}\) In Chapter 2, this condition was called “demonstrative-in-situ questions”, but in this chapter it will be called “yes/no-in-situ questions” to show the contrast with wh-in-situ questions.

\(^{13}\) As in footnote 12, this condition was called “scrambled demonstrative questions” in Chapter 2, but in this chapter it will be called “scrambled yes/no-questions” to show the contrast with scrambled wh-questions.
Table 3.1 Sample stimuli

a. Object wh-in-situ questions (wh-in-situ questions)

Ano jimotono shinbun-ni yoruto
the local newspaper-to according

その 命知らずの 冒険家が とうとう 何を 見つけたんですか。
sono inochishirazuno bokenka-ga toto nani-o mitsuketa-ndesu-ka.
the/that reckless adventurer-NOM finally what-ACC discovered- POL- Q

‘According to the local newspaper, what did that reckless adventurer finally discover?’

b. Yes/no-questions with demonstrative objects in-situ (yes/no-in-situ questions)

Ano jimotono shinbun-ni yoruto
the local newspaper-to according

その 命知らずの 冒険家が とうとう それを 見つけたんですか。
sono inochishirazuno bokenka-ga toto sore-o mitsuketa-ndesu-ka.
the/that reckless adventurer-NOM finally that-ACC discovered- POL- Q

‘According to the local newspaper, did that reckless adventurer finally discover that?’

c. Scrambled object wh-questions (scrambled wh-questions)

Ano jimotono shinbun-ni yoruto
the local newspaper-to according

何を その 命知らずの 冒険家が とうとう 見つけたんですか。
nani-o sono inochishirazuno bokenka-ga toto mitsuketa-ndesu-ka.
what-ACC the/that reckless adventurer-NOM finally discovered- POL- Q

‘According to the local newspaper, what did that reckless adventurer finally discover?’

d. Yes/no-questions with scrambled demonstrative objects (scrambled yes/no-questions)

Ano jimotono shinbun-ni yoruto
the local newspaper-to according

それを その 命知らずの 冒険家が とうとう 見つけたんですか。
sore-o sono inochishirazuno bokenka-ga toto mitsuketa-ndesu-ka.
that-ACC the/that reckless adventurer-NOM finally discovered- POL- Q

‘According to the local newspaper, did that reckless adventurer finally discover that?’
Filler items consisted of four conditions of interrogative sentences as shown in Table 3.2, namely, (a) ditransitives in canonical word order, (b) ditransitives with accusative-marked demonstrative objects scrambled within the VP, preceding the dative objects, (c) ditransitives with accusative-marked demonstrative objects scrambled within S (IP), preceding the nominative subjects, and (d) biclausal embedded clause wh-questions with wh-subjects.
Table 3.2 Filler

a. Ditransitive canonical word order

<table>
<thead>
<tr>
<th>あの出入りの</th>
<th>訪問販売員が</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ano deiri-no</td>
<td>homonhanbainin-ga</td>
</tr>
</tbody>
</table>

| その一人暮らしの老女に強引にそれを売りつけたんですか。 |
|-------------|-------------|-------------|
| sono hitorigurashi-noroo-ni goinni sore-o uritsuketa-ndesu-ka. |

‘Did the door-to-door salesman aggressively push that (product) on the lonely old woman?’

b. Ditransitive VP scrambled demonstrative objects

<table>
<thead>
<tr>
<th>あの出入りの</th>
<th>訪問販売員が</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ano deiri-no</td>
<td>homonhanbainin-ga</td>
</tr>
</tbody>
</table>

| それをその一人暮らしの老女に強引に売りつけたんですか。 |
|-------------|-------------|-------------|
| sore-o sono hitorigurashi-noroo-ni goinni uritsuketa-ndesu-ka. |

‘Did the door-to-door salesman aggressively push that (product) on the lonely old woman?’

c. Ditransitive IP scrambled demonstrative objects

<table>
<thead>
<tr>
<th>をどの出入りの</th>
<th>訪問販売員が</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sore-o ano deiri-no homonhanbainin-ga</td>
<td></td>
</tr>
</tbody>
</table>

| その一人暮らしの老女に強引に売りつけたんですか。 |
|-------------|-------------|-------------|
| sono hitorigurashi-noroo-ni goinni uritsuketa-ndesu-ka. |

‘Did the door-to-door salesman aggressively push that (product) on the lonely old woman?’

d. Biclausal embedded clause subject wh-question

| あの料理学校の講師は誰が華々しく |
|-------------|-------------|
| Ano ryorigakko-no koshi-wa dare-ga hanabanashiku |

| その本場仕込みの広東料理を披露したか話したんですか。 |
|-------------|-------------|-------------|
| sono honbajikomi-no kantonryori-o hirosita ka hanashita-ndesu-ka. |

‘Did the cooking school instructor say who showed off the authentic Cantonese cooking?’
Twenty sets (a subpart of the 200 sets used in the ERP experiment discussed below) of experimental sentences were placed in a Latin square design to create four parallel lists such that no one subject saw more than one sentence from each set. Twenty filler items (subpart of the 200 sets used in the ERP experiment discussed below) were added to each list, and then each list was pseudorandomized. 20 native speakers of Japanese (different participants from the present experiment, but the same as the participants for the reading time study discussed in Chapter 5) were asked to rate each item on a scale from 1 (odd) to 5 (natural), with ‘natural’ meaning ‘sounds natural to a native speaker, appropriate in ordinary relaxed conversation’. However, as a version of the magnitude estimation task (Stevens, 1956) that allows a more continuous and open-ended rating system than a five-point scale, subjects were instructed to imagine drawing a line whose length indicated how acceptable a given sentence was, and then to choose one of the fixed points (on the 1-to-5 scale) on a scantron that corresponded most closely to the length of their imagined line (Cowart, 1997, see Chapter 1 Section 1.2.2.1).
The results are shown in Figure 3.2. The mean acceptability ratings on the four experimental conditions ranged between 2.7 and 3.0 [wh-in-situ: 2.8 (sd 1.25); yes/no-in-situ: 3.0 (sd 1.29); scrambled wh: 2.7 (sd 1.35); scrambled yes/no: 2.8 (sd 1.27)]. Repeated measures ANOVAs were run with condition as a within-group factor, and either subject (F₁) or item (F₂) as a random factor. There was no statistically significant difference among the four conditions [omnibus ANOVA: F₁(3, 19) = 2.16, p = .093; F₂(3, 19) = 1.65, p = .178; no significantly different pairs with Tukey HSD posthoc comparisons]. Even when specific pairs were compared separately, none of the comparisons between yes/no-in-situ and scrambled yes/no-questions [F₁(1, 19) = 2.23, p = .152; F₂(1, 19) = 2.77, p = .113], wh-in-situ and scrambled wh-questions [F₁(1, 19) = .45, p = .508; F₂(1, 19) = .24, p = .631], wh-in-situ and yes/no-in-situ
questions \[F_1(1, 19) = 2.08, p = .165; F_2(1, 19) = 1.93, p = .181\], nor scrambled wh-questions and scrambled yes/no-questions \[F_1(1, 19) = .38, p = .543; F_2(1, 19) = .43, p = .518\] showed a significant or marginal difference. Therefore, there were no statistically supported differences in acceptability across conditions.

3.3 Methods

3.3.1 Subjects

Twenty (11 female) monolingual speakers of Japanese between 19 and 29 years of age (mean: 25) who had been outside of Japan for less than two years were included in the study. They were identical to the subjects who participated in the study discussed in Chapter 2. Subjects had normal or corrected-to-normal vision, were right-handed (5 reported left-handed family members), and had no neurological or reading disorders. Subjects were reimbursed for their time.

3.3.2 Materials

Stimuli consisted of four conditions of mono-clausal questions, namely, (a) wh-in-situ questions, (b) yes/no-in-situ questions, (c) scrambled wh-questions, and (d) scrambled yes/no-questions, as shown in Table 3.1 in Section 3.2 (see also Appendix 1). Scrambled conditions were included to increase the distance between wh-elements and Q-particles. 200 sets of sentences containing these four conditions were constructed. Filler items consisted of four conditions of interrogative sentences as

---

14 The total number of subjects actually run was 27. However, due to experimenter error, four subjects saw stimuli at a different presentation rate from the others, and their data had to be discarded. In addition, one subject exhibited excessive muscle tension that interfered with the EEG recording, and two subjects never returned for their second session, so their data also had to be discarded.

15 The studies discussed in Chapter 2 and the present chapter were actually run as one experiment with the same set of experimental materials and the same group of subjects. However, since the research questions addressed are different, we report the results and discussion as two separate studies.
shown in (3.14), namely, (a) ditransitives in canonical word order, (b) ditransitives with accusative-marked demonstrative objects scrambled within the VP preceding the dative objects, (c) ditransitives with accusative-marked demonstrative objects scrambled within S (IP) preceding the nominative subjects, and (d) biclausal embedded clause wh-questions with wh-subjects, as shown in Table 3.2 of Section 3.2. The in situ wh-subject *dare-ga* ‘who-NOM’ (3.14d) was used to prevent subjects from always expecting wh-words in object position (*nani-o* ‘what-ACC’ or *dare-o* ‘who-ACC’). Ditransitive sentences were used to lead subjects to expect dative objects as well. One third of the accusative objects of filler sentences were also scrambled to the beginning of the sentence (S- or IP-adjoined position), while another third were scrambled in front of the indirect object (VP-adjoined position), and the last one-third stayed in situ, the canonical position. This was done to give subjects variety in scrambling constructions so that they would not expect scrambled constituents in any particular position.

The 200 sets of stimulus sentences were placed in a Latin square design to create four parallel lists of 200 stimulus sentences such that no one subject saw more than one sentence from each set. The 200 filler sentences were added to each list, and then each list was pseudorandomized. The materials were divided into 20 sets of 20 sentences each.

### 3.3.3 Procedure

Subjects were run in two sessions lasting about 2.5 hours each. Subjects were seated in a reclining chair facing a computer monitor in a sound-attenuated room and
wore an elastic cap mounted with tin electrodes. An illuminated rectangular border appeared uninterruptedly in the middle of the screen during presentation of experimental sentences for purposes of fixation. Stimuli were presented on a computer screen in Japanese characters basically one bunsetsu at a time with 650 ms duration and 650 ms\(^{16}\) stimulus onset asynchrony. Due to experimenter error, there was no interstimulus interval between each bunsetsu\(^{17}\). A bunsetsu consists of one free morpheme (lexical word or pronoun) and the bound morpheme(s) associated with it (particles modifying the noun/verb), and will be referred to as a “word” hereafter. The interstimulus interval between sentences was three seconds, and subjects were given as much rest as they wished between sets of sentences.

In order to maintain subjects' attention, comprehension questions were inserted in the stimuli. Every five sentences on average but at a semi-random interval (at least three and at most seven comprehension questions were inserted in a set of 20 sentences), subjects were asked to answer a comprehension question regarding the immediately preceding sentence. Sentence (3.15) gives an example.

\(^{16}\) The presentation rate of 650 ms per bunsetsu is longer than the standard 500 ms per word for English, yet was decided on as the best rate after consulting three native speakers of Japanese. Similarly, Garnsey, Yamashita, Ito, and McClure (2001) used 700 ms per bunsetsu in their Japanese ERP study. Nakagome et al. (2001) used the presentation rate of 1500 ms per bunsetsu.

\(^{17}\) Because of this, sentence-final words did not show typical positive drifts; see Figures 3.5 and 3.6.
(3.15) Comprehension Question

a. Stimulus:

"あの地元の新聞によると
その 命知らずの 冒険家 とうとう 何を 見つけたんですか。"

‘According to the local newspaper, what did that reckless adventurer finally discover?’

b. Question:

"話し手は 「あの地元の新聞」に よると「その 命知らずの 冒険家」が とうとう(1.何を 2.それを)みつけたのか 関いている。

'The speaker is asking whether "the reckless adventurer" found (1. what, 2. that) according to "the local newspaper"'

Comprehension questions appeared after an interval of 1000 ms following the offset of the sentence-final word, and remained on the screen until subjects responded to them by pressing one of two buttons held in their hands. The next sentence began 3000 ms after the button press. Before beginning with the first experimental set, subjects were given a practice set of 20 sentences.

3.3.4 Electrophysiological Recording

The electroencephalogram (EEG) was recorded from 19 positions, including all standard positions of the international 10/20 system, using tin electrodes mounted
in an elastic cap. Reference electrodes were positioned on the two mastoid processes, and the EEG was algebraically referenced off-line to the mean of the activity at these two electrodes. To detect blinks and lateral eye-movements for later correction, additional electrodes were placed beneath the right eye and at the outer canthi of the two eyes. Impedances were kept below 5KΩ. The EEG was amplified with a bandpass of 0.01 to 100 Hz, digitized at 250 Hz, and stored on hard disk for off-line analysis. Data with excessive blinks were corrected using a spatial filter algorithm (Dale, 1994), and a band-pass filter set from 0.2 to 15 Hz was used on all the data prior to running analyses to reduce high frequency noise.

3.3.5 Data Analysis

Measurements were taken of both single-word averages for phasic effects, and four-word averages mid-sentence in the scrambled wh and scrambled yes/no conditions for longer-lasting effects. Single-word averages consisted of 1024 ms epochs, including a 100 ms prestimulus baseline: four-word averages consisted of 3072 ms epochs (4 x 650 ms SOA plus a 400 ms prestimulus baseline).

The statistical analyses were done separately on midline (Fz, Cz, and Pz), parasagittal (Fp1/2, F3/4, C3/4, P3/4, O1/2), and temporal (F7/8, T3/4, T5/6) electrodes. Midline analyses consisted of repeated measures ANOVAs with two within-group factors, including two levels of experimental condition type and three levels of anterior/posterior sites. Parasagittal analyses consisted of repeated measures ANOVAs with three within-group factors, including two levels of condition type, five levels of anterior/posterior sites, and two levels of hemisphere. Temporal analyses
consisted of repeated measures ANOVAs with three within-group factors, including two levels of condition type, three levels of anterior/posterior sites, and two levels of hemisphere. In addition, trend analyses were performed to investigate P200 and N350 effects (see Section 3.4 Results). An alpha level of .05 was used for all statistical tests, with a p-value of .10 considered marginally significant. The Huynh-Feld correction for lack of sphericity was applied whenever applicable. Original degrees of freedom are reported with the corrected probability level.

3.4 Results

The mean correct response rate to comprehension questions across subjects was 91% (range: 80%-97%). Thus no subject's data were excluded from the ERP analyses based on poor comprehension. The blink rejection rate (subsequently corrected by the blink-correction algorithm) was approximately 13%. The artifact rejection rate for other reasons was approximately 17%, probably due to the length of the stimulus sentences.

Recall from Table 3.1 that there were four conditions in the experimental sentences, namely, (a) wh-in-situ questions, (b) yes/no-in-situ questions, (c) scrambled wh-questions, and (d) scrambled yes/no-questions. As there were four conditions, there were six possible pairwise combinations, namely, conditions (a) vs. (b), (a) vs. (c), (a) vs. (d), (b) vs. (c), (b) vs. (d), and (c) vs. (d). Among them, two combinations are of interest here\(^\text{19}\). For effects of wh-processing, we would want to compare (a) vs. (b) or (c) vs. (d), as these combinations have the same word orders and differ only in

\(^{19}\) See Chapter 2 for the comparisons of scrambling effects.
the use of wh vs. non-wh objects. In other words, we assume that by comparing the two scrambled conditions, effects of wh-processing alone will be seen, as both conditions are equivalent in terms of scrambling. Similarly, we assume that by comparing the two in-situ conditions, only the effects of wh-processing will be seen. In the following subsections we examine the effects of wh-processing using the (a) vs. (b) and (c) vs. (d) comparisons. Sentences were compared at wh-elements to see if there was any effect of wh-processing at wh-elements themselves, and from the words following wh-elements to the sentence-final verb-Q complex to see if there were any effects following localized lexical processing of wh-elements.

3.4.1 Single-Word Averages of Wh- and Demonstrative Pronouns

As a first step, ERPs to two types of wh-pronouns (‘what-ACC' 何を and ‘who-ACC' 誰を) and two types of demonstrative pronouns (‘that-ACC' それを and ‘that-person-ACC' あのを) used in the experiment were compared, in order to see if there was any effect of wh-processing at the wh-words themselves. Visual inspection of waveforms suggested a larger P200 amplitude in response to wh-words relative to demonstratives both for the yes/no-in-situ vs. wh-in-situ comparison, as well as for the scrambled wh vs. scrambled yes/no comparison. In addition, there was a difference in the latency and amplitude of a negative peak following the P200 in both comparisons: the negative peak to the wh-words was larger and earlier than that to the demonstratives. Since the visual effects were the same across sentence positions, the averages of both types of words were collapsed across sentence positions for ease of comparison (see Figure 3.3).
Figure 3.3 ERPs (n = 20) for all electrodes to wh-words and demonstratives, collapsed across sentence positions.
To corroborate the P200 effect, ANOVAs were undertaken on the mean amplitude in a window of 200 to 300 ms poststimulus onset. There were marginal main effects in all electrode arrays [midline: \( F(1, 19) = 4.28, p = .052 \); parasagittal: \( F(1, 19) = 4.01, p = .06 \); temporal: \( F(1, 19) = 3.43, p = .08 \)], a significant interaction between condition and hemisphere in parasagittal \( F(1, 19) = 5.64, p = .028 \) and temporal \( F(1, 19) = 20.03, p < .001 \) arrays, as well as a significant interaction between condition and anteriority in the temporal \( F(2, 38) = 7.87, p = .01 \) array. These indicated that the ERPs to wh-words were more positive over right frontal regions.

Similarly, to corroborate the subsequent negative latency differences, ANOVAs were performed on the latency of the negative peak following the P200 in a window of 250 to 450 ms. There was a significant or marginal main effect in all electrode arrays [midline: \( F(1, 19) = 5.67, p = .028 \); parasagittal: \( F(1, 19) = 3.86, p = .064 \); temporal: \( F(1, 19) = 6.84, p = .017 \)]. In addition, there was a marginal interaction between condition and anteriority at the parasagittal \( F(4, 76) = 2.37, p = .079 \) electrodes. These indicated that wh-words peaked earlier than demonstratives, especially over posterior regions. Since this negative peak in response to wh-elements occurred at approximately 350 ms poststimulus onset when collapsed across electrodes and subjects, we will refer to it as the N350 hereafter. Furthermore, to corroborate the difference in amplitude between wh-elements and demonstratives in the N350 region, separate ANOVAs were performed on the mean amplitude of the latency window between 300 and 375 ms. There was a marginal or significant main
effect in all electrode arrays \textit{[midline]}: F(1, 19) = 4.30, p = .052; \textit{parasagittal}: F(1, 19) = 3.64, p = .072; \textit{temporal}: F(1, 19) = 5.66, p = .028], as well as a significant condition \times hemisphere interaction at the parasagittal [F(1, 19) = 6.90, p = .017] and temporal [F(1, 19) = 11.58, p = .003] arrays, indicating that wh-words were more negative than demonstratives, especially over the left hemisphere.

These statistical results with the P200 and N350, comparing wh-elements to demonstratives, initially gave the impression of a wh-effect. However, a closer examination of four different pronominal elements (see Figure 3.4) suggested that this was merely an effect of lexical processing differences. Table 3.3 summarizes the properties of the four pronouns used for the following analyses.
Figure 3.4 ERPs (n = 20) at a left frontal electrode (F3) to four types of pronominal elements, collapsed across sentence positions.

Table 3.3 Characteristics of four pronoun types collapsed across electrodes and subjects

<table>
<thead>
<tr>
<th>Pronouns</th>
<th>Gloss</th>
<th>Orthography</th>
<th>Stroke Count</th>
<th>Graphic Complexity</th>
<th>Writing Conventionality</th>
<th>Length</th>
<th>P200 Amplitude (µV)</th>
<th>N350 Latency (ms)</th>
<th>N350 Amplitude (µV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dare-o</td>
<td>'who-ACC'</td>
<td>誰を</td>
<td>18</td>
<td>9</td>
<td>4.7%</td>
<td>2</td>
<td>2.258</td>
<td>354.46</td>
<td>-0.357</td>
</tr>
<tr>
<td>Nani-o</td>
<td>'what-ACC'</td>
<td>何を</td>
<td>10</td>
<td>5</td>
<td>100.0%</td>
<td>2</td>
<td>1.492</td>
<td>344.52</td>
<td>-0.593</td>
</tr>
<tr>
<td>Ano-hito-o</td>
<td>'that-person-ACC'</td>
<td>あの人に</td>
<td>9</td>
<td>3</td>
<td>100.0%</td>
<td>4</td>
<td>1.805</td>
<td>368.77</td>
<td>0.014</td>
</tr>
<tr>
<td>Sore-o</td>
<td>'that-ACC'</td>
<td>それを</td>
<td>6</td>
<td>1.5</td>
<td>99.9%</td>
<td>3</td>
<td>1.090</td>
<td>360.76</td>
<td>0.054</td>
</tr>
</tbody>
</table>
As for the P200 effect, along with general claims that relate it to visual targets and attention (Luck & Hillyard 1994; O'Donnell, Swearer, Lloyd, & Hokama 1997, among others), it was noticed that wh-words involve characters more complex than pronouns, and this may have caused the differences in P200 amplitude shown in Figures 3.3 and 3.4. As shown in Table 3.3, wh-words include a higher ratio of kanji, symbolic Chinese characters, that tend to be graphically more complex and require more strokes to write. In addition, it was also noticed that in the experimental sentences, ‘that’, ‘that-person’, and ‘what’ were written with conventional combinations of kanji and kana (syllabaries), while ‘who’ was written with a rather non-conventional kanji alternative. A corpus count of the Asahi Shinbun (Amano & Kondo 2000), a popular Japanese newspaper, confirmed the discrepancy. More than 99% of the first three words were written in the way shown in Table 3.3, while only 4.7% of the occurrences of ‘who’ were written with kanji, as 誰, instead of with kana, as だれ. In accord with this observation, the P200 was largest in response to ‘who’ (誰).

Trend analyses were used to determine if the P200 amplitude, between 200 and 300 ms post-stimulus onset collapsed over electrode site for the four pronominal elements, was influenced by the above factors in a linear fashion. Supporting the assumed linear correlation, the linear component for total stroke count was significant $[F(1, 57) = 14.21, p < .001]$, whereas none of the other higher order (quadratic, cubic) trend components were significant. On the other hand, although the linear component for graphic complexity (average number of strokes per character: total stroke count /
number of characters) was also significant \( F(1, 57) = 10.10, p = .002 \), so was the cubic component \( F(1, 57) = 5.24, p = .016 \). Since cognitive hypotheses rarely extend beyond predicting a linear relationship, we do not have any reason to believe that the higher order trends are meaningful. When other higher order trends as well as linear trends are significant, it may simply indicate random variation in the data (C. Van Petten, personal communication). Therefore, we concluded that graphic complexity did not show a strict linear correlation with P200 amplitude, given the significant higher order trend. Similarly, for writing conventionality (percentage of a particular \textit{kanji-kana} combination out of the total occurrences in the corpus), the linear trend component \( F(1, 57) = 11.09, p = .002 \) as well as the quadratic trend component \( F(1, 57) = 5.25, p = .028 \) were significant, which disproves a monotonic increase in amplitude with declining writing conventionality. The correlation between P200 amplitude and word length (character count) was also tested, yet did not reveal a significant linear trend component. Thus the larger P200 in response to wh-words is suggested to reflect stroke count.

As for the N350 effect, in light of claims that correlate the negative peak latency following the P200 with frequency and word length in lexical processing paradigms (Osterhout, Bersick, & McKinnon, 1997; King & Kutas, 1998), trend analyses were used to determine if N350 peak latency, between 250 and 450 ms post-stimulus onset collapsed over electrodes for the four pronouns, was influenced by frequency or word length. While frequency did not, word length did reveal a significant linear trend component \( F(1, 57) = 6.89, p = .012 \), with no significant
higher trend components. As further possibilities, as in the analysis of the P200 effect, the orthogonal trends between N350 latency and stroke count, graphic complexity, and writing conventionality were tested, yet none of the linear trends was significant.

As for the amplitude difference of the N350, mean amplitudes between 300 and 375 ms poststimulus onset collapsed across electrodes for the four pronominal elements were tested for a linear trend with the possible factors examined above. Among stroke count, graphic complexity, writing conventionality, and word length, word length was the only factor that revealed a significant linear trend component alone \( \text{linear: } F(1, 57) = 4.76, p = .033 \); \( \text{quadratic: } F(1, 57) = 1.38, p = .246 \); \( \text{cubic: } F(1, 57) = .71, p = .403 \).

From the above, we concluded that the larger P200 in response to wh-words reflects stroke count, whereas the N350 effect for wh-words, both in terms of peak latency and mean amplitude, has something to do with word length (see Section 3.5 Discussion for more details).

3.4.2 Between Wh and Verb-Q: Wh vs. Demonstrative Sentences

As a second step following the wh-word comparison, ERPs to words following wh-words up to and including the verb-Q (question particle) complex were examined. The two scrambled conditions and the two in-situ conditions ((b) vs. (d) and (a) vs. (c) in Table 3.1) were compared separately, as it was assumed that comparing conditions identical in word order would reveal the effects of wh-processing alone.
3.4.2.1 Scrambled Wh- vs. Yes/No-Questions

The gloss of the relevant experimental sentences with the point of comparison capitalized is given below in (3.16).

(3.16)

a. The local newspaper-to according

\begin{align*}
&\text{what-ACC that RECKLESS ADVENTURER-NOM FINALLY \_ DISCOVERED-Q.} \\
&\text{‘According to the local newspaper, what did that reckless adventurer finally discover?’}
\end{align*}

b. The local newspaper-to according

\begin{align*}
&\text{that-ACC that RECKLESS ADVENTURER-NOM FINALLY \_ DISCOVERED-Q.} \\
&\text{‘According the local newspaper, did that reckless adventurer finally discover that?’}
\end{align*}

As seen in (3.16), both sentence types had five words after the scrambled objects: a subject noun phrase with three words, an adverbial, and a verb (‘that reckless adventurer-NOM finally discovered-Q’ in the sample stimuli in (3.16)). However, since wh-phrases showed P200 and N350 differences relative to demonstratives, as discussed above, the ERPs to ‘what-ACC’ and ‘that-ACC’ were not identical enough to be used as feasible pre-stimulus baselines. Instead, the demonstrative sono ‘that’ of the subject noun was used as the pre-stimulus baseline. This yielded four-word averages of an adjective, a noun, an adverbial, and a verb-Q complex, as in ‘reckless adventurer-NOM finally discovered-Q’, capitalized in (3.16).

Visual inspection of the four-word averages showed a right-lateralized negativity at about 500 ms following the onset of ‘finally’ to the sentence end of the scrambled wh-question condition (see Figure 3.5).
Figure 3.5 ERPs to words following wh-words up to and including the verb-question particle complex. (a) ERPs (n = 20) at midline and parasagittal electrodes (Fp1/2, F3/4, C3/4, P3/4, O1/2, Fz, Cz, Pz) relative to a 400 ms prestimulus baseline. Shown are the ERPs covering the adjective, subject noun, adverbial, and verb-question particle positions ('reckless adventurer-NOM finally discovered-Q') of scrambled wh- vs. yes/no-questions.
To capture the first observation, ANOVAs were performed in the latency window of 500 ms poststimulus onset of ‘finally’ to the sentence end including the sentence final verb-Q complex ‘discovered-Q’ (1800 to 2600 ms from the baseline). A significant main effect of condition was found in all three arrays [midline: $F(1, 19) = 6.66$, $p = .018$; parasagittal: $F(1, 19) = 6.19$, $p = .022$; temporal: $F(1, 19) = 5.26$, $p = .033$]. In addition, there was a marginal interaction between condition and anteriority in the midline array [$F(2, 38) = 3.13$, $p = .088$], as well as a significant interaction between condition and hemisphere in the parasagittal array [$F(1, 19) = 8.78$, $p < .001$]. These interactions were due to greater negativity over right anterior regions in response to scrambled wh-sentences. To summarize, scrambled wh-questions in comparison to scrambled yes/no-questions elicited a greater negativity, especially over the right hemisphere, both at the adverbial and at the verb-Q complex positions.

### 3.4.2.2 Wh-in-Situ vs. Yes/No-in-Situ Questions

The relevant stimuli in gloss with the point of comparison capitalized are given in (3.17).

(3.17)

a. The local newspaper-to according

that reckless adventurer-nom finally **what-ACC** DISCOVERED-Q.

‘According to the local newspaper, what did that reckless adventurer finally discover?’

b. The local newspaper-to according

that reckless adventurer-NOM finally **that-ACC** DISCOVERED-Q.

‘According the local newspaper, did that reckless adventurer finally discover that?’
In this comparison, the sentence-final verb ('discovered-Q') was the only word following in-situ wh-words or demonstratives, thus measurements were undertaken on single-word averages of the brain responses to the sentence-final verb-Q complex.

Similar to the comparison of the scrambled conditions discussed above, wh-in-situ questions appeared to elicit greater negativity 300 ms after the onset of ‘discovered-Q’, larger over the right hemisphere (see Figure 3.6).
**Figure 3.6.** ERPs (n = 20) for all electrodes relative to a 100 ms prestimulus baseline. Shown are the ERPs to the final verb-Q position (`discovered-Q`) of wh-in-situ vs. yes/no-in-situ-questions.
This observation was in part statistically supported by ANOVAs performed within the latency window of 300-600 ms.\textsuperscript{20} Only the temporal array showed a significant main effect \[F(1, 19) = 4.69, p = .043\] and a marginal condition x hemisphere interaction \[F(1, 19) = 3.90, p = .063\] that indicated a greater right-lateralized negativity in response to wh-in-situ questions compared to yes/no-in-situ questions. In addition, although not statistically significant, the negativity to wh-in-situ questions appeared to be more posterior in its distribution compared to scrambled wh-questions.

3.4.3 Summary of Effects

To summarize, wh-words elicited larger P200s and a subsequent delayed, larger amplitude negative peak in the latency range of 250 to 450 ms (N350) compared to demonstratives. The P200 seemed to be related to stroke count, whereas both the latency and the amplitude of the N350 seemed to be related to word length. As for the scrambled wh vs. yes/no comparison, scrambled wh-questions elicited greater right-lateralized negativity at the adverbial and verb-Q positions. Wh-in-situ questions compared to yes/no-in-situ questions elicited greater right-lateralized negativity at the sentence-final verb-Q position as well, although the statistical results were weaker.

\textsuperscript{20} The results were basically the same even when the time window was extended to 300-900 ms. Only the temporal array showed a significant main effect \[F(1, 19) = 4.69, p = .031\] and a marginal condition x hemisphere interaction \[F(1, 19) = 3.40, p = .081\] that indicated a greater right-lateralized negativity in response to wh-in-situ questions compared to yes/no-in-situ questions.
3.5 Discussion

3.5.1 Lexical Effects on Wh-Words

Recall that wh-words elicited greater P200s and a later negative peak with a larger amplitude in the latency range of 250 to 450 ms (N350) compared to demonstratives (see Figures 3.3 and 3.4). The P200 effect was correlated with stroke count while the N350 effect was correlated with word length.

The interaction between P200 amplitude and stroke count appears to reflect local visual processing of complex Chinese characters. This is consistent with reports of larger P200s when local features are attended to (Kotchoubey, Wascher, & Verleger, 1997), as well as with general claims relating the P200 to visual targets and attention (Luck & Hillyard, 1994; O’Donnell et al., 1997). Related to this, Takashima, Ohta, Matsushima, and Toru (2001) reported that Japanese content words (nouns, verbs) elicit larger P200s than Japanese function words (particles, auxiliary verbs). However, this effect is confounded by the fact that content words tend to be written in Chinese characters (kanji), while function words are usually written in visually less complex phonemic characters (kana). The results, showing that even graphically complex closed-class words elicit larger P200s, strongly suggest that visual complexity rather than word class is the crucial variable in P200 amplitude.

Turning now to the N350 effect, Neville, Mills, and Lawson (1992) reported that closed-class words elicited a negative peak around 280 ms at left frontal regions (N280), whereas open-class words elicited a negative peak around 350 ms at temporal and posterior regions (N350). In a reply, Osterhout, Bersick, and McKinnon (1997)
argued that the negative peak latencies at frontal sites were actually correlated with word length and frequency, rather than with word class itself. They reported that peak latency was highly correlated with both mean length (the longer, the later) and mean log frequency (the less frequent, the later) of words. Similarly, King and Kutas (1998) reported a correlation between word frequency and peak latency. They suggested that what Neville et al. (1992) termed the N350 and N280 might actually be something else. The former is actually the N400, which predominantly occurs at right posterior regions and varies in amplitude. The latter is a separate ERP component they termed the frequency sensitive negativity (FSN), which occurs predominantly at left anterior regions and varies in peak latency according to frequency (the more frequent, the earlier). They did not find any significant correlation between this component and word class or length.

In the present study, word length (but not frequency, stroke count, graphic complexity, or writing conventionality) was the only factor that showed a significant linear trend for both the later peak latency and the larger amplitude of the N350 effect in responses to wh-words compared to demonstratives. This goes along with Osterhout et al.’s (1997) claim, in that longer demonstratives elicited later peaks than shorter wh-words. As for the peak amplitude difference, Neville et al. (1992) reported that among open class words, longer words elicited more positive ERPs than shorter words following the P200 over frontal regions (Neville et al. 1992: Figure 5). Although the morphology of their data differs from that of our data, in both cases, longer words (demonstratives in the present study) elicited more positive (or less
negative) ERPs than shorter words (wh-words in the present study) after the P200. It is interesting that frequency, which has been reported to exert strong influence over lexical ERP effects in English (Osterhout et al., 1997; King & Kutas, 1998), played no apparent role in the present study.

We initially presumed that since wh-words and demonstratives were both closed-class words, they would provide a reasonable basis for investigating wh-processing effects, yet the comparison revealed only orthogonal lexical processing effects of little interest to our original concern. However, they do provide valuable additional evidence in the debate over open-vs.-closed-class lexical processing differences. In particular, our findings support the view that P200 (in Japanese) and N350 (in Japanese and English) lexical processing differences are not a reflection of word class difference, but of extraneous differences related to visual processing alone, namely graphic complexity and length.

3.5.2 Effect of Wh between Wh and Verb-Q Complex

Recall that wh-conditions in comparison to yes/no-conditions elicited (a) right-lateralized negativity at the adverbial and verb-Q positions in scrambled wh-questions (see Figure 3.5) and (b) right lateralized negativity at the verb-Q (300-600 ms) position in wh-in-situ questions (see Figure 3.6).

As can be seen by comparing Figures 3.5 and 3.6, the negative difference at the sentence-final position showed a statistically stronger amplitude difference in scrambled wh-questions than in wh-in-situ questions. Moreover, in the scrambled comparison, the negativity had rather anterior distribution, yielding a significant
condition x anteriority interaction at the midline electrodes in response to the verb-Q complex. On the other hand, in the in-situ comparison, the negativity had a visually posterior distribution, although this never reached statistical significance. Thus one may wonder whether or not these two effects are the same or not. Despite the difference in amplitude and anteriority, the two effects have several features in common. One is condition: both types of effects occurred in response to wh-questions in comparison to yes/no-questions. The others are point of occurrence (after processing wh-elements), polarity (negative), and lateralization (right-lateralized).

From the above, it seems plausible to consider the two effects at least related to each other. Therefore, in this section, we provisionally treat the two as one type of right-lateralized negativity effect and proceed with our discussion on what it might index.

The following subsections will discuss four alternative interpretations of this effect.

**3.5.2.1 Interpretation 1: Wh-Q Dependency Effect?**

Recall that in the introduction, Section 3.1, we predicted that there might be a dependency between Japanese wh-elements and their Q-particles similar to filler-gap dependencies in English or German, and that such a dependency might be indexed by LAN or slow anterior negative potentials, as repeated here in (3.18).
(3.18)

a. Filler-gap dependency for wh-movement languages (e.g., English, German)

What did Calvin bring __?
FILLER            GAP
|___________|  
(L)AN

b. Wh-Q dependency for wh-in-situ languages? (e.g., Japanese)

カービンが何を持ってきたんですか。
Calvin-ga nani-o mottekita-ndesu-ka
Calvin-NOM what-ACC brought-POLITE-Q
WH Q
|____________|  
(L)AN?

‘What did Calvin bring?’

In the present experiment, we found right-lateralized negativity at the adverbial and verb-Q positions in scrambled wh-questions and at the verb-Q position in wh-in-situ questions. This may be evidence for a dependency between Japanese wh-elements and their related Q-particles in terms of the parser’s expectation for Q-particles (Miyamoto & Takahashi, 2001, 2002b; Aoshima et al., in press). In particular, the negativity started at the adverbial position preceding the verb-Q position in scrambled wh-sentences. One could claim that subjects were eagerly anticipating the verb-Q at the adverbial position, which caused the negativity there. The negativity was less robust (most statistical results were marginal) when the wh-elements were in-situ and right next to the verb-Q. This may be due to the fact that subjects did not have to wait very long for the Q-particle, in accordance with Gibson’s (1998, 2000) model of parser preference in minimizing expectancy.
In addition, although we did not see any P600 effect in the present experiment, the wh-Q dependency hypothesis might also be related to the findings of Nakagome et al. (2001), who showed that wh-questions with a [-Q] particle elicit a P600 effect. This could indicate that since subjects could not find the Q-particle they had expected after seeing the wh-word in the Nakagome et al. experiment, they may have had to reformulate some sort of syntactic expectation (similar to other P600 responses to garden-paths; see Osterhout and Holcomb (1992) for an early example), thus causing the P600 effect.

Therefore, the first part of the hypothesis, that there is a dependency between a Japanese wh-element and its related Q-particle, seems to be well supported. However, the second part, that the wh-Q dependency is similar to a filler-gap dependency, needs further consideration. First, the slow negative anterior potentials reported in response to filler-gap dependencies in English and German (King & Kutas, 1995; Kluender & Münte, 1998; Fiebach et al., 2001) are long-lasting, whereas the effects in the present experiment were not. For the in-situ comparison, we only had the sentence final verb to compare, and for the scrambled comparison, the effect did not start until the adverbial position. Second, the reported LAN effects (Kluender & Kutas 1993 a,b; King & Kutas, 1995; Kluender & Münte, 1998; Fiebach et al., 2001) are either bilateral or left-lateralized, but our effects were right-lateralized. However, the present study is not the only case of right-lateralized negativity. Müller, King, and Kutas (1997) used an auditory version of King and Kutas (1995) and reported right-lateralized anterior negativity to object relative clause sentences compared to subject
relative clause sentences. The authors speculated that this was probably due to some mechanism of auditory perception, such as more involvement of the right hemisphere. We will return to these latency and lateralization issues in Chapter 4.

3.5.2.2 Interpretation 2: Wh-Scope Calculation Effect?

Recall further that in Section 3.1 (Introduction) we predicted that there might be an interpretive process involved in calculating the scope of Japanese wh-elements, and that we might see such an effect at the verb-Q position that disambiguates it, as repeated here in (3.19).

\[(3.19)\]
\[
\text{scope-calculation effect?} \\
\text{LAN?, P600??} \\
\downarrow \\
\text{カルビンが 何を 持ってきたんですか。} \\
\text{Calvin-NOM nani-o mottekita-ndesu-ka} \\
\text{WH what-ACC brought-POLITE-Q} \\
\text{‘What did Calvin bring?’}
\]

Again recall that in the present experiment, we found a right-lateralized negativity at the adverbial and verb-Q positions in scrambled wh-questions and at the verb-Q position in wh-in-situ questions. This could have been due to subjects calculating the scope of the wh-elements at the verb-Q position. Subjects might have first realized what the scope of the wh-element they had previously seen was when viewing the Q-particle, and this process might have elicited the negativity.

The drawback to this interpretation is the sentence position, although it can better explain why the negativity in scrambled wh-questions started at the adverbial position. However, subjects might have thought that the sentence end (and the
attached Q-particle) had come already at the adverbial position. Since 75% of the experimental materials (stimuli and filler combined) were mono-clausal, subjects might well have expected a mono-clausal structure and a forthcoming verb-Q when they saw an adverbial and started preemptive scope-calculation. We will return to this interpretation in Chapter 4.

3.5.2.3 Interpretation 3: Prosody Effect?

Another possible explanation for the right-lateralized negativity is a prosody effect. The difference between wh and yes/no-questions can be indicated by different intonation contours at the ends of spoken sentences (the critical word in the present paradigm), and this could have caused the ERP effects. Although stimuli were presented visually in the present experiment, subjects may have been saying sentences to themselves silently, and this may have caused a tacit intonation effect. We explore this possibility below.

Beyond the lexical prosody level, there are said to be two levels of prosodic structure in Japanese called accentual phrase and intonational phrase21 (Beckman & Pierrehumbert, 1986; Beckman, 1996). The accentual phrase is defined by its tonal markings, and it is the domain of two delimitative peripheral tones, the phrasal H (high) and the boundary L (low). An accentual phrase is composed of one to three words. The intonational phrase is the domain for downstep (or catathesis), which is the compression and lowering of a pitch range following an accented phrase. An intonational phrase generally measures between one and three accentual phrases. Within an intonational phrase, a series of lexical accents are compressed in the form of

---

21 In Beckman & Pierrehumbert (1986), this level is called “intermediate phrase”.

a descending staircase. Then the pitch is reset upwards at the beginning of each new intonational phrase.

Japanese wh-elements such as *nani* ‘what’ and *dare* ‘who’ are accented in wh-questions\(^{22}\) and are said to form an intonational phrase with the following word(s), resulting in their deaccentuation or pitch reduction (Maekawa, 1991a, b)\(^{23,24}\). Thus, when a wh-element and a verb are next to each other and the verb is lexically accented, as in *nani-ga mieru* ‘what-NOM is.seen’, the accent on the verb disappears. On the other hand, the accent on the verb remains for yes/no-questions, as in *nani-ka mieru* ‘something is.seen’ (see Figure 3.7).

![Figure 3.7](image)

**Figure 3.7** The F\(_0\) (fundamental frequency) contours of wh-question *nani-ga mieru* ‘what-NOM is.seen’ (left) and yes/no-question *nanika mieru* ‘something is.seen’ (right) as uttered by a male Tokyo Japanese speaker. (adopted from Maekawa, 1991a: 203)

---

\(^{22}\) They are not accented in quantifier contexts, such as *nanika* ‘something’, *nanimo* ‘nothing’, and *nandemo* ‘everything’.

\(^{23}\) We thank Mary Beckman for pointing us to Maekawa’s work.

\(^{24}\) Maekawa (1991a,b) argues that this is a part of the general correlation between prosodic salience and focus, in that focused elements in a sentence (wh-element in a wh-question) call for an increased pitch and the subsequent pitch reduction of the following utterances. See also Deguchi and Kitagawa (2002).
It is possible that subjects may have imagined wh-induced pitch reduction when they read wh-questions but not when they read yes/no-questions, and this may have influenced their brain responses. However, since stimuli were presented word by word instead of as entire sentences, it is questionable whether subjects would have created intonational phrases across individual presentations of words. One native speaker of Japanese we consulted felt that each presentation of a word became its own intonation unit. On the other hand, there have been reports of the influence of implicit prosodic structure in ambiguity resolution with word-by-word self-paced reading times (Bader, 1998; Hirose, 1999; Fodor, 2002).

Thus assuming the possibility of subjects creating intonational phrases across presentations of words, a prosody analysis of sample tokens was conducted. Two native speakers of Japanese (one male and one female, Tokyo dialect) recorded 24 experimental sentences each (6 sets x 4 conditions) that were pseudorandomized with 24 filler items. The digitized speech signals (44.1 kHz/16 bit sampling rate) were analyzed in terms of F0 (fundamental frequency) contours. The F0 maxima of the verb (before the sentence final LH contour for interrogatives) for all conditions and the subject noun for the scrambled conditions were measured and compared using paired one-tailed t-tests. Table 3.4 shows the F0 maxima of the relevant words, and Figure 3.8 shows typical F0 contours for the four experimental conditions. As for wh-in-situ (a) in comparison to yes/no-in-situ (b) questions, a salient peak at ‘what/who-ACC’ and a reduced pitch at the following verb position [male speaker: t(5) = 3.29, p
= .011; female speaker: t(5) = 1.90, p = .065] were observed. As for scrambled wh-, (c), in comparison to scrambled yes/no-, (d), questions, a salient peak at ‘what/who-ACC’ and a reduced pitch starting from the following word [(at subject noun)\textsuperscript{25} male speaker: t(5) = 4.81, p = .002; female speaker: t(5) = 5.03, p = .004] to the sentence-end [male speaker: t(5) = 3.22, p = .012; female speaker: t(5) = 2.78, p = .003] were observed.

Table 3.4 Mean F0 maxima (in Hz) of relevant words

<table>
<thead>
<tr>
<th></th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Subject</td>
<td>Verb</td>
</tr>
<tr>
<td></td>
<td>Noun</td>
<td></td>
</tr>
<tr>
<td>a. Wh-in-Situ</td>
<td>NA</td>
<td>92.0</td>
</tr>
<tr>
<td>b. Yes/No-in-Situ</td>
<td>NA</td>
<td>115.3</td>
</tr>
<tr>
<td>c. Scrambled Wh</td>
<td>113.8</td>
<td>97.2</td>
</tr>
<tr>
<td>d. Scrambled Yes/No</td>
<td>152.5</td>
<td>115.0</td>
</tr>
</tbody>
</table>

\textsuperscript{25} The word immediately following, \textit{sono} ‘the’, was integrated into the accentual phrase with the scrambled wh-/non-wh-object and was not easily measurable independently.
Figure 3.8 The F0 contours of the last six words of (a) wh-in-situ question, (b) yes/no-in-situ question, (c) scrambled wh-question, and (d) scrambled yes/no-question as uttered by a male Tokyo Japanese speaker.
Based on the above observations, we cannot deny the possibility that wh-in-situ-questions have a different intonation pattern from yes/no-in-situ-questions at the verb-Q position, and that this could have resulted in the right-lateralized negativity effect. However, the pitch reduction with scrambled wh-questions was long-lasting: it lasted from right after the wh-words were first uttered until the sentence-end. If the right-lateralized negativity effect was due to pitch reduction induced by a wh-word, we would have expected to see it start right after the wh-words. However, we saw the effect only for the two sentence-final words, which renders the intonation account for the scrambled comparison questionable at best. Furthermore, if both in-situ and scrambled comparisons yielded the right-lateralized negativity effect, it seems more plausible to have one factor (such as Q-dependency or scope-calculation) to account for both comparisons rather than intonation for the in-situ comparison and something else for the scrambled comparison.

3.5.2.4 Interpretation 4: End-of-Sentence/Dispreferred Parse Effect?

Another possible interpretation for the right-anterior negativity effect is an end-of-sentence wrap-up effect. N400-like negativities were reported by Osterhout and Holcomb (1992), Hagoort, Brown, and Groothusen (1993), and Osterhout and

---

26 Another possibility is that at the end of the sentence, subjects repeated to themselves the entire sentence with an appropriate intonation contour, and this may have caused the sentence final right-lateralized negativity effect for both in-situ and scrambled comparisons. However, this interpretation is questionable for two reasons. First, recall that the sentence-final effects started within 300 ms of the final verb-Q complex. It is doubtful whether subjects could recall the entire sentence that rapidly. Second, even if we assume that they could be really fast at repeating stimulus sentences subvocally, it is doubtful whether they could read them only once and memorize them well enough to repeat them at the end.

27 As for prosody-related ERP effects, there has been a report of an N400-P600 pattern reflecting a prosody-induced garden-path effect and a large positive waveform at intonational phrase boundaries termed “closure positive shift” (CPS) (Steinhauer, Alter, & Friederici, 1999).
Nicol (1999) at the end of sentences containing anomalies. They interpreted such a negativity as a sentence-final-wrap-up effect. More recently, Friederici and Frisch (2000) reported a widely-distributed sentence-end negativity with sentences also containing anomalies. However, there are two points that are different in the present study. First, unlike those studies with anomalies, the stimulus sentences used in the present study were all grammatical. Additionally, while the above mentioned sentence-end negativity effects had either posterior or wide-spread distribution, the distribution of the effects found in the present study was anterior at least for the scrambled wh- v. yes/no-question comparison. These points seem to diminish the possibility of a sentence-end wrap-up effect.

Another possibility is a dispreferred parse effect, in that the effects were due to a grammatical but dispreferred sentence construction. However, dispreferred parse effects with grammatical sentences have been linked to the P600 (Osterhout & Holcomb, 1992; Osterhout, Holcomb, & Swinney, 1994; Friederici, Mecklinger, & Hahne, 1996) instead of to any negativity to our best knowledge. For instance, among the sentence types with dispreferred constructions, sentence-end negativity was found only in response to ungrammatical conditions (Osterhout et al., 1994). In other words, grammatical dispreferred parses seem to result in P600 effects at sentence-

---

28 Kluender and Kutas (1993b) also reported that a posterior N400-like negativity effect at the end of wh-island violation sentences, such as *What do you wonder who they caught __ at __ by ACCIDENT?, and wh-if sentences, such as ?What do you wonder if they caught him at __ by ACCIDENT?. Interestingly, only half of their 30 subjects showed the N400 effect, while the other half showed a P600 effect. Kluender and Kutas attribute the sentence-final N400/P600 effect to an overloaded sentence-final-wrap-up effect for filler-gap assignment.
intermediate positions and no apparent sentence-final negativities\textsuperscript{29}, making the right-lateralized negativity effects in the present study unlikely to be a dispreferred effect.

\subsection*{3.5.2.5 Wh-Q Dependency or Wh-Scope Calculation?}

To summarize thus far, we found an ERP effect (right-lateralized (anterior) negativity) to wh-questions at the verb-Q position. We have reviewed several possible ways of interpreting this result. Among them, the prosody, sentence-end wrap-up, and dispreferred parse accounts do not seem to be as plausible as an account involving some type of relationship between a wh-element and a Q-particle.

The question, however, is whether the relationship indicates an effect of the parser’s expectation for a Q-particle, or more of the local processing cost of wh-scope calculation at a verb-Q position. The present findings, in conjunction with the findings of prior studies (Nakagome et al., 2001; Miyamoto & Takahashi 2001, 2002b; Aoshima et al., in press), seem to favor to the first interpretation, although the evidence is not yet conclusive. It should be noted that these two interpretations are not mutually exclusive. Subjects may expect a Q-particle as soon as they see a wh-element, but this may be due to the need for wh-scope calculation. It is hard to tease apart the two possible effects using the mono-clausal design of this experiment, which points to the need for more data from a multi-clausal design. Therefore, a subsequent experiment on the processing of bi-clausal wh-questions focused more closely on these issues (see Chapter 4).

\textsuperscript{29} In Osterhout and Holcomb (1992), sentences like \textit{The broker hoped to sell the stock was sent to jail} in comparison to sentences like \textit{The broker persuaded to sell the stock was sent to jail} elicited an N400-like effect at the sentence-final word. The first group of sentences received a 14\% acceptance rate while the second group of sentences received a 80\% acceptance rate. We consider that a 14\% acceptance rate does not really constitute a grammatical sentence.
3.6 Conclusion

In summary, Japanese wh-questions in comparison to yes/no-questions elicited two types of effects. One type involved lexical effects among wh-words, namely enhanced P200 and N350 components. The P200 effect correlated with stroke count, while the N350 effect, both in terms of length and amplitude, was due to word length. The other type of effect involved right-lateralized negativities between wh-words and verb-Q complexes. This appeared to be due to some dependency between wh-elements and a Q-particle, possibly involving the parser's expectation for a Q-particle or clausal scope calculation, or both. The experiment reported in Chapter 4 on the processing of bi-clausal wh-questions was designed to address these issues.
Chapter 4: ERP Measures of Bi-Clausal Japanese Wh-Questions

4.1 Introduction

4.1.1 Remaining Questions from the Mono-Clausal Experiment (Chapter 3)

This chapter examines the ERP processing of bi-clausal Japanese wh-questions. Recall from Chapter 3 that mono-clausal wh-questions elicited right-lateralized, negativity as shown in (4.1).

(4.1)

卡尔ビンが 何を 持ってきたんですか。
Calvin-ga nani-o mottekita-ndesu-ka

‘What did Calvin bring?’

For scrambled wh-questions, the effect had an anterior distribution and was seen in response to the final two words of the sentence, namely, the adverbial and the verb-Q. For wh-in-situ questions, the effect had a wide-spread distribution and was seen in response to the sentence-final verb-Q position.

The remaining question was what this right (anterior) negativity effect indexes in terms of language processing and our two hypotheses, namely, the wh-Q dependency hypothesis and the wh-scope calculation hypothesis. Recall that the wh-Q dependency hypothesis predicts that subjects expect a Q[uestion]-particle as soon as they see a wh-element, and that right (anterior) negativity indexes the effect of
subjects’ expectation for a Q-particle. On the other hand, the wh-scope calculation hypothesis predicts that there is an interpretive process involved in calculating the scope of wh-elements and that right (anterior) negativity indexes the effect of wh-scope calculation. The data obtained in Chapter 3 can be interpreted to be consistent with both, and we were not able to determine which hypothesis was right.

4.1.2 Present Experiment: Bi-clausal design

In order to focus more on this question, the present study examined the processing of bi-clausal wh-questions involving embedded and matrix clause scope, which were expected to involve various types of scope calculation-related operations. Stimuli were slightly modified versions of Miyamoto and Takahashi’s (2001, 2002b) reading time study, as shown in (4.2).

(4.2)

a. Embedded clause wh-questions

専務が どんな パソコンを 買ったか　経理の 係長が　聞きましたか。
[senmu-ga donna pasokon-o katta-ka] keiri-no kakaricho-ga__kikimashita-ka?
director-N what.kind.of PC-A bought-Q accounting-of manager-N asked.POL-Q

‘Did the accounting manager ask what kind of computer the director bought?’

b. Matrix clause wh-questions

専務が どんな パソコンを 買ったと　経理の 係長が　言いましたか。
[senmu-ga donna pasokon-o katta-to] keiri-no kakaricho-ga__iimashita-ka?
director-N what.kind.of PC-A bought-that accounting-of manager-N said.POL-Q

‘What kind of computer did the accounting manager say the director bought?’
c. Embedded clause yes/no-questions

専務が新しいパソコンを買ったか経理の係長が聞きましたか。

‘Did the accounting manager ask if the director bought a new computer?’

d. Matrix clause yes/no-questions

専務が新しいパソコンを買ったと経理の係長が言いましたか。

‘Did the accounting manager say the director bought a new computer?’

In this stimulus design, the embedded clause was scrambled to the beginning of the sentence, which is typical for a long embedded clause in Japanese (Yamashita, 2002).

Unlike Miyamoto and Takahashi’s (2001, 2002b) experiment, the embedded object NP *donna pasokon-o* ‘what.kind.of PC-ACC’ was presented as two separate *bunsetsu* in order to increase the temporal distance between wh-words and the embedded verb-Q complex and to be consistent with word (*bunsetsu*) by word presentation. Similarly, a modifier (*keiri-no* ‘accounting-of’ in (4.2)) of the matrix subject was added to extend the distance between the embedded verb and the matrix verb-Q complex.

With respect to other, more minor modifications, recall that in Chapter 3 we saw lexical effects in response to wh-words, namely larger amplitude P200 and N350 components correlated with stroke count and word length. In order to see if there was any effect of wh-processing beyond these lexical aspects, all of the wh and non-wh modifiers were matched for word length. Half of the stimulus pairs contained *donna* (どんな) ‘what.kind.of’ as the wh-modifier and an adjective such as *atarashii* (新しい) ‘new’ as the non-wh counterpart. These pairs were matched for number of
characters, but had different Chinese character (kanji) to phonemic character (kana) ratios (e.g., *donna どんな* ‘what.kind.of’ consists of all kanas, while *atarashii 新しい* ‘new’ consists of one kanji and two kanas). The other half of the stimuli contained numeral classifiers such as *nan-bai-no (何杯の)* ‘what-cup-of’ as the wh-modifier and numeral classifiers such as *ni-hai-no (二杯の)* ‘2-cup-of’ as the non-wh counterpart. These wh- and non-wh-classifiers were written in essentially the same way except for one character, *nan (何)* ‘what’ vs. *ni ‘2’ (二)*, which would minimize stroke count effects. In addition, the embedded verbs were all in the simple past tense and the matrix verbs were in the neutral polite form *ii-mashi-ta-ka ‘said-POL-PAST-Q’*, instead of the casual form *i-tta-no ‘say-PAST-Q’, to better adhere to conventions of written presentation.

4.1.3 Predictions

The predicted results are illustrated below in (4.3).
(4.3) Predictions

a. Embedded clause wh-questions

Wh-Q RAN?

Wh-Scope RAN?

専務が どんな パソコンを 買ったか 経理の 係長が 聞きましたか。
[senmu-ga donna pasokon-o katta-ka] keiri-no kakaricho-ga_kikimashita-ka?
director-N what.kind.ofPC-A bought-Q accounting-of manager-N asked.POL-Q

‘Did the accounting manager ask what kind of computer the director bought?’

b. Matrix clause wh-questions

Wh-Q RAN?

Wh-Scope RAN?

専務が どんな パソコンを 買ったと 経理の 係長が 言いましたか。
[senmu-ga donna pasokon-o katta-to] keiri-no kakaricho-ga__iimashita-ka?
director-N what.kind.ofPC-A bought-that accounting-of manager-N said.POL-Q

‘What kind of computer did the accounting manager say the director bought?’

If the wh-Q dependency hypothesis is correct and right (anterior) negativity (hereafter called “RAN”) indexes the effect of the parser’s on-going expectation for a Q-particle, we should see it between the wh-element and embedded verb-Q positions in embedded clause wh-questions (4.3a), and between the wh-element and the matrix verb-Q in matrix clause wh-questions (4.3b). If the wh-scope calculation hypothesis is correct and RAN indexes the effect of the parser’s local calculation of wh-scope, then we should see it where wh-scope becomes disambiguated, namely, at the embedded verb-Q position of embedded clause wh-questions (4.3a), and at the matrix verb-Q position of matrix clause wh-questions (4.3b).
One crucial comparison was the embedded verb, ‘bought-Q/that’ in (4.3). If RAN indexes the parser’s expectation for a Q-particle, the effect should be elicited by matrix clause wh-questions in comparison to embedded clause wh-questions, as there should be an unresolved dependency at the embedded verb-‘that’ complex of matrix clause wh-questions. On the other hand, if RAN indexes an effect of wh-scope calculation, we should see it in response to embedded clause wh-questions in comparison to matrix wh-questions, as there should be a scope calculation operation for embedded clause wh-questions (but not matrix clause wh-questions) at the embedded verb-Q position. In addition, we should be able to test if RAN is really a sentence-end wrap-up effect (although this is not very likely due to the grammaticality of the stimuli and the anterior distribution of the negativity, as discussed in Chapter 3). With a bi-clausal design, we should be able to elicit RAN at sentence positions other than at sentence end.

4.2 Norming

A paper-and-pencil rating study on a subset of the experimental items was conducted in order to norm the experimental materials. Stimuli consisted of four conditions of bi-clausal questions, namely, (a) embedded clause wh-questions, (b) matrix clause wh-questions, (c) embedded clause yes/no-questions, and (d) matrix clause yes/no-questions, as shown in Table 4.1. The embedded clause was scrambled (displaced) to the beginning of the sentence (gap position indicated by underscoring), which is common for a heavy object in Japanese (Yamashita & Chang, 2001; Yamashita, 2002).
Table 4.1 Sample stimuli

a. Embedded clause wh-questions

専務が どんな パソコンを 買ったか 経理の 係長が 聞きましたか。

\[ \text{senmu-ga } \text{donna pasokon-o katta-ka} \]  \text{keiri-no kakaricho-ga } \text{kikimashita-ka.} \\
\text{director-N what.kind.of PC-A bought-Q accounting-of manager-N asked.POL-Q}

‘Did the accounting manager ask what kind of computer the director bought?’

b. Matrix clause wh-questions

専務が どんな パソコンを 買ったと 経理の 係長が 言いましたか。

\[ \text{senmu-ga } \text{donna pasokon-o katta-to} \]  \text{keiri-no kakaricho-ga } \text{iimashita-ka.} \\
\text{director-N what.kind.of PC-A bought-that accounting-of manager-N said.POL-Q}

‘What kind of computer did the accounting manager say the director bought?’

c. Embedded clause yes/no-questions

専務が 新しい パソコンを 買ったか 経理の 係長が 聞きましたか。

\[ \text{senmu-ga } \text{atarashii pasokon-o katta-ka} \]  \text{keiri-no kakaricho-ga } \text{kikimashita-ka.} \\
\text{director-N new PC-A bought-Q accounting-of manager-N asked.POL-Q}

‘Did the accounting manager ask if the director bought a new computer?’

d. Matrix clause yes/no-questions

専務が 新しい パソコンを 買ったと 経理の 係長が 言いましたか。

\[ \text{senmu-ga } \text{atarashii pasokon-o katta-to} \]  \text{keiri-no kakaricho-ga } \text{iimashita-ka.} \\
\text{director-N new PC-A bought-that accounting-of manager-N said.POL-Q}

‘Did the accounting manager say the director bought a new computer?’
Filler material consisted of four conditions of declarative/interrogative sentences as shown in Table 4.2, namely, (a) bi-clausal declaratives, (b) mono-clausal ditransitive declaratives, (c) mono-clausal ditransitive yes/no-questions, and (d) mono-clausal ditransitive wh-questions (see also Appendix 2).
Table 4.2 Filler

a. Bi-clausal declaratives

bijutsukan-no kancho-ga shisanka-ga tasu-no kaiga-o kifushita-to nobe-mashi-ta.
museum-of director-N millionaire-N many-of picture-A donated-that state-POL-PAST

'The director of the museum said that the millionaire donated many pictures’

b. Mono-clausal ditransitive declaratives

hanaya-no tenin-ga onen-no daijoyu-ni hanataba-o todoke-mashi-ta.
flower.shop-of clerk-N past.year-of big.actress-D bouquet-A deliver-POL-PAST

'The clerk of the flower shop delivered a bouquet to the former movie star actress’

c. Mono-clausal ditransitive yes/no questions

beteran-no hodouin-wa seifuku-no shojo-o hogosha-ni watashi-mashi-ta-ka.
veteran-of school.counselor-T school.uniform-of girl-A guardian-D hand-POL-PAST-Q

'Did the veteran school counselor hand over a girl in her school uniform to her guardian?’

d. Mono-clausal ditransitive wh-questions

zaikuna shachou.fujin-wa nan-chaku-no doresu-o kokyuten-ni chumonshi-mashi-ta-ka.
high-end president.wife-T what-piece-of dress-A boutique-D order-POL-PAST-Q

'How many dresses did the high-end wife of the president place an order of to the boutique?'
24 sets (a subset of the 200 sets used in the ERP experiment discussed below) of experimental sentences were placed in a Latin square design to create four parallel lists such that no one subject saw more than one sentence from each set. 24 filler items were added to each list, and each list was pseudorandomized. 20 native speakers of Japanese (different participants from the present experiment, but the same as the participants for the reading time study discussed in Chapter 5) were asked to rate each item on a scale from 1 (odd) to 5 (natural), with ‘natural’ meaning ‘sounds natural to a native speaker, appropriate in ordinary relaxed conversation’. However, as a version of the magnitude estimation task (Stevens, 1956) that allows a more continuous and open-ended scaling system than a five-point scale, subjects were instructed to imagine drawing a line whose length indicated how acceptable a given sentence was, and then to choose one of the fixed points (on a 1-to-5 scale) on a scantron that corresponded most closely to the length of their imagined line (Cowart, 1997, see Chapter 1 Section 1.2.2.1).
The results are shown in Figure 4.1. The mean acceptability ratings of the four experimental conditions ranged from 3.0 to 3.6 [embedded-wh: 3.6 (sd 1.29); matrix-wh: 3.0 (sd 1.23); embedded yes/no: 3.3 (sd 1.22); matrix yes/no: 3.3 (sd 1.31)]. Repeated measures ANOVAs were run with condition as a within-group factor, and either subject (F₁) or item (F₂) as a random factor. There was a statistically significant difference by subjects and a marginal difference by items across the four conditions [omnibus ANOVA: F₁(3, 19) = 6.30, p = .001; F₂(3, 23) = 2.34, p = .075]. Tukey HSD posthoc comparisons revealed that embedded clause wh-questions were significantly more acceptable than matrix clause wh-questions for both subject and item analyses. Separate pair-wise comparisons again showed that embedded clause wh-questions were significantly more acceptable than matrix clause wh-questions in

**Figure 4.1** Mean acceptability ratings of experimental materials.
both subject and item analyses \[ F_1(1, 19) = 13.48, p = .002; F_2(1, 23) = 6.71, p = .015, \]
and that embedded clause wh-questions were significantly more acceptable than
embedded clause yes/no-questions in the subject-wise analysis \[ F_1(1, 19) = 11.85, p = .003; F_2(1, 23) = 2.14, p = .15 \] \(^1\). The pairwise comparisons between matrix clause
wh- and matrix clause yes/no-questions \[ F_1(1, 19) = 2.66, p = .119; F_2(1, 23) = 2.65, p = .115 \] and between embedded clause yes/no- and matrix clause yes/no-questions
\[ F_1(1, 19) = .00, p = 1.000; F_2(1, 23) = .01, p = .919 \] revealed no significant or
marginal difference.

4.3 Methods

4.3.1 Subjects

Twenty (13 female) monolingual speakers of Japanese between 19-29 years of
age (mean: 23) who had been outside Japan for less than two years were included in
the study. \(^2\) They were a different group of Japanese speakers from those who
participated in the studies discussed in Chapters 2 and 3. Subjects had normal or
corrected-to-normal vision, were right-handed (4 reported left-handed family
members), and had no neurological or reading disorders. Subjects were reimbursed
for their time.

\(^1\) This difference in acceptability was generally correlated with the amplitudes of right anterior
negativity effects we will see later in the Chapter. However, the total relationship between acceptability
ratings and on-line measures (both ERPs and reading times) for the bi-clausal wh-questions will be
discussed in Chapter 5.

\(^2\) The total number of subjects actually run was 22. However, one subject showed low comprehension
scores, and there was an equipment malfunction during the session of another subject. Thus their data
were discarded.
4.3.2 Materials

Stimuli consisted of four conditions of bi-clausal questions, namely, (a) embedded clause wh-questions, (b) matrix clause wh-questions, (c) embedded clause yes/no-questions, and (d) matrix clause yes/no-questions, as shown in Table 4.1 in Section 4.2 (see also Appendix 2). 200 sets of sentences containing these four conditions were constructed. Matrix verbs for (a) and (c) (‘say’ verbs) and for (b) and (d) (‘ask’ verbs) were matched for character count and kanji ratio. The token frequency of the two groups of the verbs did not differ significantly. Filler items consisted of four conditions of declarative or interrogative sentences, namely, (a) bi-clausal declaratives, (b) mono-clausal ditransitive declaratives, (c) mono-clausal ditransitive yes/no-questions, and (d) mono-clausal ditransitive wh-questions as shown in Table 4.2 in Section 4.2. Declarative sentences were used to prevent subjects from always expecting interrogative endings. Similarly, mono clausal sentences were used to prevent subjects from always expecting bi-clausal construction. Ditransitive sentences were used to lead subjects to expect dative objects as well.

The 200 sets of stimulus sentences were placed in a Latin square design to create four parallel lists of 200 stimulus sentences such that no one subject saw more than one sentence from each set. The 150 filler sentences were added to each list, and

3 A corpus count of Asahi Shinbun (Amano & Kondo 2000), a popular Japanese newspaper, was used to examine the token frequency of the verbs used. Some of the verbs had the token frequency of their verb forms listed, while others only the token frequency of their noun forms only (e.g., v. tennis-suru ‘tennis-do’ --> n. tennis ‘tennis’, as Japanese has many ‘noun-do’ verbs). Only the pairs that had the token frequency of either both verb forms or both noun forms listed were considered, yielding 169 pairs out of the 200 pairs. Paired t-test comparing the token frequency of the two groups of verbs [mean token frequency: 30,817 (sd 40,094) for ‘say’ verbs; 29,727 (sd 29,708) for ‘ask’ verbs] showed no significant difference between the two groups [(168) = -.374, p = .709].
then each list was pseudorandomized. The materials were divided into 10 sets of 35 sentences each.

**4.3.3 Procedure**

Subjects were run in one session lasting about 2.5 hours. Subjects were seated in a recliner facing a computer monitor in a sound-attenuated room, and wore an elastic cap mounted with tin electrodes. An illuminated rectangular border appeared uninterruptedly in the middle of the screen during presentation of experimental sentences for purposes of fixation. Stimuli were presented on a computer screen in Japanese characters basically one *bunsetsu* at a time with 450 ms duration and 650 ms stimulus onset asynchrony.\(^4\) A *bunsetsu* consists of one free morpheme (lexical word or pronoun) and the bound morpheme/s associated with it (particles modifying the noun/verb), and will be referred as a “word” hereafter. The interstimulus interval between sentences was three seconds, and subjects were given as much rest as they wished between sets of sentences.

In order to maintain subjects’ attention, comprehension questions were inserted in the stimuli. Every seven sentences on average but at a semi-random interval (at least three and at most six comprehension questions were inserted in a set of 35 sentences), subjects were asked to answer a comprehension question regarding the immediately preceding sentence. (4.4) gives an example. A verb followed by two nouns was presented, and subjects had to decide which noun was the subject of the

---

\(^4\) This ISI of 200 ms between *bunsetsu* was deemed optimal after consulting five native speakers of Japanese.
verb in the preceding sentence. Subjects were asked to respond to the questions by pressing one of two buttons held in their hands.

(4.4) Comprehension Question

a. Stimulus:

専務がどんなパソコンを買ったか経理係長が聞きましたか。

[senmu-ga donna pasokon-o katta-ka] keiri-no kakaricho-ga=kikimashita-ka?

director-N what.kind.ofPC-A bought-Q accounting-of manager-N asked.POL-Q

‘Did the accounting manager ask what kind of computer the director bought?’

b. Question:

聞く1.専務2.係長

kiku senmu kakaricho

ask director manager

Before beginning with the first experimental set, subjects were given a practice set of 20 sentences.

4.3.4 Electrophysiological Recording

The electroencephalogram (EEG) was recorded from 19 positions, including all standard positions of the international 10/20 system, using tin electrodes mounted in an elastic cap. Reference electrodes were positioned on the two mastoid processes, and the EEG was algebraically referenced off-line to the mean of the activity at these two electrodes. To detect blinks and lateral eye-movements for later correction, additional electrodes were placed beneath the right eye and at the outer canthi of the two eyes. Impedances were kept below 5KΩ. The EEG was amplified with a

---

5 This task was adopted from Nagata (1993), and was the same method used by Miyamoto and Takahashi (2001).

6 The response hand was not counterbalanced to control for dominance, yet this is not expected to yield a problem, as the accuracy but not the speed of the response was recorded.
bandpass of 0.01 to 100 Hz, digitized at 250 Hz, and stored on hard disk for off-line analysis. Data with excessive blinks were corrected using a spatial filter algorithm (Dale, 1994), and a band-pass filter set from 0.2 to 15 Hz was used on all the data prior to running analyses to reduce high frequency noise.

4.3.5 Data Analysis

Measurements were taken both of single and two-word averages for phasic effects, and of four-word averages for longer-lasting effects. Single-word averages consisted of 1024 ms epochs including a 100 ms prestimulus baseline, and two-word averages consisted of 2048 ms epochs including a 200 ms prestimulus baseline. Four-word averages consisted of 3072 ms epochs (4 x 650 ms SOA plus a 400 ms prestimulus baseline).

The statistical analyses were done separately on midline (Fz, Cz, and Pz), parasagittal (Fp1/2, F3/4, C3/4, P3/4, O1/2), and temporal (F7/8, T3/4, T5/6) electrodes. Midline analyses consisted of repeated measures ANOVAs with two within-group factors, including two levels of experimental condition type and three levels of anterior/posterior sites. Parasagittal analyses consisted of repeated measures ANOVAs with three within-group factors, including two levels of condition type, five levels of anterior/posterior sites, and two levels of hemisphere. Temporal analyses consisted of repeated measures ANOVAs with three within-group factors, including two levels of condition type, three levels of anterior/posterior sites, and two levels of hemisphere. An alpha level of .05 was used for all statistical tests, with a p-value of .10 considered marginally significant. The Huynh-Feld correction for lack of
sphericity was applied whenever applicable. Original degrees of freedom are reported with the corrected probability level.\textsuperscript{7}

4.4 Results

The mean correct response rate to comprehension questions across subjects was 87\% (range: 70\%-100\%). The blink rejection rate (subsequently corrected by the blink-correction algorithm) was approximately 18\%. The artifact rejection rate for other reasons was approximately 4\%. Comparisons were made to examine the two hypotheses discussed in the Introduction (i.e., the wh-Q dependency hypothesis and the wh-scope calculation hypothesis).

4.4.1 Wh-Scope Calculation Hypothesis

Recall that the wh-scope calculation hypothesis predicts that if RAN indexes local wh-scope calculation, it should be seen at the embedded verb-Q position in embedded clause wh-questions, and at the matrix verb-Q position in matrix clause wh-

\textsuperscript{7} We ran overall ANOVAs on the four experimental conditions in Chapter 2, as it was feasible to compare across all four conditions. However, for the present study, specific comparisons were preplanned. Due to this, we did not run omnibus ANOVAs, a choice justified by statisticians such as G. Keppel:

When “tests designed to shed light on these particular questions are planned before the start of an experiment, ...a researcher is not interested in the omnibus F test—this test is more appropriate in the absence of specific hypotheses...analytical comparisons can be conducted directly on a set of data without reference to the significance or non-significance of the omnibus F test.” (Keppel 1982: 106)

Following a modified Bonferroni procedure (Keppel, 1982) that lets up to \lbrack number of conditions -1\rbrack contrasts to maintain the same alpha level, an alpha level of .05 was also used for planned pairwise comparisons.

\textsuperscript{8} There were three subjects who scored in seventies, at 78\%, 72\%, and 70\%; the other 17 subjects scored between 84\% and 100\%, making their mean score 90\%. Two of the three subjects who scored in the seventies were male, and male subjects were harder to find than female subjects. There was an equipment malfunction during the experimental session of the 22nd subject, which at the time of data analysis was not restored. Considering the above, the data of the three subjects were included.
questions. This hypothesis was tested by measuring ERPs at the embedded verb and matrix verb positions.

4.4.1.1 At the Embedded Verb

The gloss of the relevant experimental sentences with the point of comparison capitalized is given below in (4.5).

(4.5)

a. [Director-N what.kind.of PC-A BOUGHT-Q] accounting-of manager-N asked-Q
   ‘Did the accounting manager ask what kind of computer the director bought?’

b. [Director-N what.kind.of PC-A BOUGHT-THAT] accounting-of manager-N said-Q
   ‘What kind of computer did the accounting manager ask the director bought?’

c. [Director-N what.kind.of PC-A BOUGHT-Q] accounting-of manager-N asked-Q
   ‘Did the accounting manager ask if the director bought a new computer?’

At this position, only the embedded clause wh condition (4.5a) was expected to show the RAN effect according to the wh-scope calculation hypothesis, as it was the only condition that underwent wh-scope calculation at the scope-disambiguating Q-particle. It was expected that scope calculation for the matrix clause wh condition (4.5b) should not occur until the matrix verb-Q position, and that there should be no wh-scope calculation for the yes/no control condition (4.5c) at all. Thus comparisons were made between embedded vs. matrix clause wh-questions and between embedded clause wh- vs. yes/no-questions.

Figure 4.2 shows ERPs to embedded vs. matrix clause wh-questions. Contrary to predictions of the wh-scope calculation hypothesis, a RAN effect was seen in
response to matrix clause wh-questions rather than to embedded clause wh-questions. ANOVAs performed in the 300-600 ms window showed no significant or marginal differences, but ANOVAs in the 300-900 ms window revealed a marginal condition x anteriority x hemisphere interaction in the temporal array \( F(2, 38) = 2.88, p = .071 \), indicating that matrix clause wh-questions in comparison to embedded clause wh-questions were more negative especially over the right anterior regions of scalp. Although there was a visible N400-like effect to embedded clause wh-questions at posterior electrodes, ANOVAs run in the window of 350-450 ms revealed no significant or marginal differences.
専務がどんなパソコンを買ったか 経理の係長が聞きましたか。
[Director-N what.kind.of PC-A bought-Q] accounting-of manager-N asked-Q

Figure 4.2 ERPs (n=20) recorded from all electrodes relative to a 100 ms prestimulus baseline. Shown are the ERPs to the embedded verb position (‘bought-Q/that’) of embedded clause wh vs. matrix clause wh questions.
Figure 4.3 shows ERPs to embedded clause wh- vs. yes/no-questions. Visual inspection indicated no apparent RAN effect to either embedded clause wh- or yes/no-questions. ANOVAs performed in the 300-600 ms and 300-900 ms windows showed no significant or marginal effects. Surprisingly, on visual inspection there again appeared to be an N400-like effect at this sentence position, this time however, in response to embedded yes/no-questions relative to embedded wh-questions. However, ANOVAs run in the window of 350-450 ms revealed no significant or marginal differences.

---

9 Due to a concern that there may have been some effect of forming an indirect question in embedded clause yes/no-questions (as in ‘The accounting manager asked if the director bought a new computer’), ERPs to embedded clause wh-questions were also compared to ERPs to matrix clause yes/no-questions. However, visual inspection showed no observable effect, and ANOVAs run in the 300-600 ms and 300-900 ms windows showed no significant or marginal differences.
……… Embedded Wh

専務がどんなパソコンを買ったか　経理の係長が聞きましたか。
[Director-N what.kind.of PC-A bought-Q] accounting-of manager-N _ asked-Q

Embedded Yes/No

専務が新しいパソコンを買ったか　経理の係長が聞きましたか。
[Director-N new PC-A bought-Q] accounting-of manager-N _ asked-Q

Figure 4.3 ERPs (n=20) recorded from all electrodes relative to a 100 ms prestimulus baseline. Shown are the ERPs to the embedded verb position (‘bought-Q’) of embedded clause wh- vs. yes/no-questions.
To summarize, contrary to predictions of the wh-scope calculation hypothesis, RAN was not seen in response to embedded clause wh-questions relative to matrix wh-questions at the embedded verb position; it was instead seen in response to matrix clause wh-questions although statistically marginally. Comparisons of embedded clause yes/no-questions with embedded wh-questions revealed no effects.

4.4.1.2 At the Matrix Verb

The gloss of the relevant experimental sentences with the point of comparison capitalized is given below in (4.6).

(4.6)

a. [Director-N what.kind.of PC-A bought-that] accounting-of manager-N SAID-Q
   ‘What kind of computer did the accounting manager say the director bought?’

b. [Director-N what.kind.of PC-A bought-Q] accounting-of manager-N ASKED-Q
   ‘Did the accounting manager say what kind of computer the director bought?’

c. [Director-N what.kind.of PC-A bought-that] accounting-of manager-N SAID-Q
   ‘Did the accounting manager say that the director bought a new computer?’

At this position, according to the wh-scope calculation hypothesis, only the matrix clause wh condition (4.6a) was expected to elicit a RAN effect, as it was the only condition that underwent wh-scope calculation at the sentence-final scope-disambiguating Q-particle. It was presumed that the scope calculation for the embedded clause wh condition (4.6b) should already have been completed at the embedded verb-Q position, and that no wh-scope calculation at all should have been required for the yes/no control condition (4.6c). Thus comparisons were made
between embedded vs. matrix clause wh-questions and between matrix clause wh- vs. yes/no-questions.

Figure 4.4 shows ERPs to matrix clause and embedded clause wh-questions. There was no observable RAN effect in this comparison, and ANOVAs performed in the 300-600 ms and 300-900 ms windows showed no significant or marginal differences between the two conditions. However, increased negativity around 300-400 ms was observed in response to matrix clause wh-questions. ANOVAs on mean amplitude and peak amplitude in the 300-400 ms window revealed a marginal or significant condition x anteriority interaction in the midline array [mean amplitude: F(2, 38) = 4.79, p = .039; peak amplitude: F(2, 38) = 3.05, p = .093], indicating that matrix clause wh-questions were more negative at anterior regions.
Figure 4.4 ERPs (n=20) recorded from all electrodes relative to a 100 ms prestimulus baseline. Shown are the ERPs to the matrix verb position ('asked/said-Q') of embedded vs. matrix wh-questions.
Figure 4.5 shows ERPs to matrix clause wh- vs. yes/no-questions. There was no observable RAN effect, and ANOVAs performed in the 300-900 ms window showed no significant or marginal differences between the two conditions. ANOVAs run in the time window of 300-600 ms showed a marginal condition x hemisphere x anteriority interaction in the parasagittal array \( [F(4, 76) = 2.12, p = .089]\). However, this was because matrix clause yes/no-questions were more negative than matrix clause wh-questions especially at left central sites. More generally, on visual inspection, matrix clause yes/no-questions showed an N400-like effect. ANOVAs run in the latency window of 350-450 ms showed a marginal main effect of condition in the midline \( [F(1, 19) = 3.60, p = .073]\) and parasagittal \( [F(1, 19) = 3.09, p = .095]\) arrays, indicating a larger N400 peak for matrix clause yes/no-questions compared to matrix clause wh-questions. Additionally, there was a significant condition x hemisphere interaction in parasagittal \( [F(1, 19) = 5.56, p = .029]\) and temporal \( [F(1, 19) = 5.50, p = .03]\) arrays, as well as a significant condition x hemisphere x anteriority interaction in the parasagittal array \( [F(4, 76) = 4.28, p = .009]\). These interactions showed that matrix clause yes/no-questions were more negative than matrix clause wh-questions at left central regions (Figure 4.6). However, although the effect was left-lateralized, the absolute values of the negative peak in response to matrix clause yes/no-questions were right-lateralized (Figure 4.7), following the standard distribution of the N400. Thus the difference seemed have been due to greater left-right asymmetry in the matrix wh-question condition.
Figure 4.5 ERPs (n=20) recorded from all electrodes relative to a 100 ms prestimulus baseline. Shown are the ERPs to the matrix verb position (‘asked-Q’ in Table 4.1) of matrix wh- vs. yes/no-questions.
**Figure 4.6** Mean amplitude N400 responses for matrix clause wh vs. yes/no-questions at the matrix verb-Q position for parasagittal electrodes.

**Figure 4.7** Mean amplitude N400 responses for matrix clause wh vs. yes/no-questions at the matrix verb-Q position. Mean amplitudes are given averaged across right vs. left electrodes of the parasagittal and temporal arrays.
In order to further investigate this N400 effect, matrix clause yes/no-questions were compared to embedded clause-yes/no-questions (Figure 4.8). N100 and N400 effects were observed in response to matrix clause yes/no-questions. ANOVAs run in the window of 50-150 ms revealed a marginal main effect of condition in the midline array [F(1, 19) = 3.02, p = .098], indicating a larger N100 to matrix clause yes/no-questions. Separate ANOVAs in the 350-450 ms window showed a significant main effect in all three arrays [midline: F(1, 19) = 7.37, p = .014; parasagittal: F(1, 19) = 7.48, p = .013; temporal: F(1, 19) = 5.74, p = .027], supporting the observation of an N400 effect.
専務が新しいパソコンを買ったか経理の係長が聞いたか。

専務が新しいパソコンを買ったと経理の係長が言いましたか。

---

**Figure 4.8** ERPs (n=20) recorded from all electrodes relative to a 100 ms prestimulus baseline. Shown are the ERPs to the matrix verb position (‘asked/said-Q’) of embedded vs. matrix yes/no-questions.
In summary, similar to the comparisons at the embedded verb-Q position, no comparison at the matrix verb-Q position showed RAN effects that would have been predicted by the wh-scope calculation hypothesis. The negative peak in response to matrix clause wh-questions compared to embedded clause wh-questions was larger at anterior regions (Figure 4.4), but this difference had a very short latency window (within 100 ms), unlike other RAN effects that last more than 300 ms. For the matrix clause wh vs. yes/no comparison, an N400-like effect was found to matrix yes/no-questions (Figure 4.5). We suspect that this is related to the presumed surprise at seeing a Q-particle at the end of a sentence that could otherwise have been declarative; recall that filler material consisted of 75 declarative sentences (see Section 4.5 Discussion for more details). Together with comparisons at the embedded verb position, comparisons at the matrix verb position do not support the wh-scope calculation hypothesis.

### 4.4.2 Wh-Q Dependency Hypothesis

The wh-Q dependency hypothesis predicts that if RAN indexes subjects’ expectation for a Q-particle, it should be seen between wh-elements and Q-particles in both embedded and matrix clause wh-questions. This hypothesis was tested by investigating ERPs between wh-elements and Q-particles. The following subsections examine wh vs. yes/no comparisons at various sentence positions.

#### 4.4.2.1 At the Wh-Modifier

Recall that in 50% of the stimuli, the wh-modifier to the embedded object noun (‘what.kind.of PC-ACC’ in Table 4.1) was *donna* ‘what.kind.of’ with the non-wh-
counterpart being an adjective such as *atarashii* ‘new’ (‘new PC-ACC’ in Table 4.1).

The other 50% of the stimuli contained numeral classifiers such as *nan-bai-no* ‘what-cup-of’ for the wh-conditions or *ni-hai-no* ‘2-cup-of’ for the non-wh-conditions. ERPs to wh- vs. non-wh-modifiers were compared in order to examine if there was any effect of Q-expectation on wh-modifiers following the wh-Q dependency hypothesis.

Figure 4.9 shows ERPs to all wh-modifiers vs. non-wh-modifiers. An anterior negativity effect in response to wh-modifiers, as well as an N400 effect to non-wh-modifiers was observed. As for the anterior negativity, ANOVAs run in the window of 300-900 ms\(^{10}\) showed a marginal main effect of condition \([F(1, 19) = 3.06, p = .096]\) and a significant condition x anteriority interaction in the parasagittal \([F(4, 76) = 3.99, p = .027]\) and temporal \([F(2, 38) = 4.68, p = .04]\) arrays. These results indicated that wh-modifiers were more negative than non-wh-modifiers especially in anterior regions. As for the N400, ANOVAs in the window of 350-450ms revealed a significant condition x anteriority interaction at the parasagittal \([F(4, 76) = 5.32, p = .001]\) and temporal \([F(2, 38) = 6.46, p = .018]\) arrays, indicating that non-wh-modifiers had a larger N400 at posterior sites.

\(^{10}\) Due to the N400 effect to non-wh-modifiers, the 300-600 ms window was not considered to be a good latency window to test the anterior negativity effect to wh-modifiers. The ANOVAs run in this time window showed a significant condition x anteriority interaction in the parasagittal \([F(4, 76) = 3.66, p = .035]\) and temporal \([F(2, 38) = 4.31, p = .048]\) arrays, indicating that non-wh-modifiers were more negative than wh-words at posterior sites, while wh-words were more negative than non-wh-words at anterior sites. This visually appeared to be due to the overlap with the N400 effect to non-wh words and the anterior negativity effect to wh-words.
Figure 4.9 ERPs (n=20) recorded from all electrodes relative to a 100 ms prestimulus baseline. Shown are the ERPs to all wh- vs. non-wh-modifiers.
Figure 4.10 shows ERPs to ‘what.kind.of’ and the adjective modifiers such as *atarashii* ‘new’. An anterior negativity effect to ‘what.kind.of’, as well as P200 and N400 effects to adjectives were observed. In order to corroborate the anterior negativity effect, ANOVAs were run in the latency window of 300-900 ms\(^{11}\). There was a significant or marginal condition x anteriority interaction in all arrays [*midline*: \(F(2, 38) = 3.1, p = .09\); *parasagittal*: \(F(4, 76) = 4.49, p = .018\); *temporal*: \(F(2, 38) = 8.8, p = .001\)], indicating that ‘what.kind.of’ was more negative in response to wh-phrases over anterior regions.

As for the P200 effect to adjectives, ANOVAs run in the 200-300ms window revealed a significant or marginal main effect of condition in all three arrays [*midline*: \(F(1, 19) = 5.29, p = .033\); *parasagittal*: \(F(1, 19) = 5.03, p = .037\); *temporal*: \(F(1, 19) = 3.86, p = .064\)], as well as a significant condition x anteriority interaction in the parasagittal \(F(4, 76) = 3.77, p = .031\) and temporal \(F(2, 38) = 8.99, p = .005\) arrays. These results indicated that adjectives had a larger P200 than ‘what.kind.of’, especially over anterior regions.

\(^{11}\) Due to the N400 effect to adjectives, the 300-600 ms window was not considered to be a good latency window to test the anterior negativity effect to ‘what.kind.of’. The ANOVAs run in this time window showed a significant or marginal condition x anteriority interaction in all arrays [*midline*: \(F(2, 38) = 3.36, p = .08\); *parasagittal*: \(F(4, 76) = 4.97, p = .013\); *temporal*: \(F(2, 38) = 9.99, p = .004\)], indicating that non-wh-modifiers were more negative than wh-words at posterior sites (and that wh-words were more negative than non-wh-words at anterior sites for parasagittal and temporal arrays). Visually, this appeared to be due to the overlap with the N400 effect to non-wh words and the anterior negativity effect to wh-words.
Figure 4.10 ERPs (n=20) recorded from all electrodes relative to a 100 ms prestimulus baseline. Shown are the ERPs to ‘what.kind.of’ vs. adjectival modifiers.
As for the N400 effect to adjectives, separate ANOVAs were run in the window of 350-450 ms. There was a significant or marginal main effect of condition 

\[\text{midline: } F(1, 19) = 10.34, p = .005; \text{ parasagittal: } F(1, 19) = 4.85, p = .04; \text{ temporal: } F(1, 19) = 3.79, p = .067\], as well as a significant or marginal condition x anteriority interaction 

\[\text{midline: } F(2, 38) = 4.2, p = .052; \text{ parasagittal: } F(4, 76) = 8.61, p = .001; \text{ temporal: } F(2, 38) = 15.17, p = .001\] in all three arrays. These results indicated that the N400 to adjectives was larger than to ‘what.kind.of’, especially over posterior sites.

Figure 4.11 shows ERPs to the numeral classifier pairs, as in ‘what-cup-of’ (how many cups of)’ vs. ‘2-cup-of’. A right-lateralized negativity effect was observed in response to the wh-condition. ANOVAs run in the window of 300-600 ms showed a significant main effect of condition at the parasagittal array \[F(1, 19) = 5.24, p = .034\] and a marginal condition x hemisphere interaction at the temporal array \[F(1, 19) = 4.03, p = .059\], indicating that ‘what-cup-of’ is more negative especially over the right hemisphere. Additionally, ANOVAs run in the window of 300-900 ms showed a significant main effect of condition in all three arrays 

\[\text{midline: } F(1, 19) = 4.5, p = .047; \text{ parasagittal: } F(1, 19) = 7.68, p = .012; \text{ temporal: } F(1, 19) = 6.14, p = .023\], again indicating that ‘what-cup-of’ is more negative than ‘2-cup-of’.

Furthermore, a small P200 effect to wh-modifiers in comparison to wh-modifiers (this time the opposite of the ‘what.kind.of’ vs. ‘new’ comparison that yielded a P200 effect to non-wh-modifiers) was observed, yet ANOVAs run in the 200-300ms window revealed no significant or marginal main effect.
Figure 4.11 ERPs (n=20) recorded from all electrodes relative to a 100 ms prestimulus baseline. Shown are the ERPs to wh-numeral ‘what.kind.of’ vs. non-wh-numeral ‘2-cup-of’ classifiers.
To summarize, ERPs to wh-modifiers were more negative in comparison to non-wh modifiers in the latency window of 300-900 ms (and also in the window of 300-600 ms anteriorly). Collapsing across all wh- vs. non-wh-modifier pairings revealed an effect of anterior negativity in response to wh-modifiers and an N400 effect to non-wh-modifiers. The comparison between *donna* ‘what.kind.of’ and adjectives revealed anterior negativity in response to wh-modifiers, as well as larger amplitude P200 and N400 components in response to adjectives. The P200 and N400 effects seem to be related to the graphic complexity and lower frequency of adjectives compared to ‘what kind of’ (see Section 4.5 Discussion), as they disappeared when the graphic complexity and frequency of wh- vs. non-wh-modifiers were more closely controlled for in the numeral classifier comparison. For the classifier comparison, there was only a right-lateralized negativity effect. The N400 effect found in the collapsed wh- vs. non-wh-modifier comparison thus seems to be due to the 50% of non-wh-modifiers that were adjectives.

**4.4.2.2 Between Wh and Embedded Verb**

The gloss of the relevant experimental sentences with the point of comparison capitalized is given below in (4.7).
Both types of (i.e., both embedded clause and matrix clause) wh-questions and yes/no-questions were collapsed, as they were identical up to the embedded verb position.

Figure 4.12 shows ERPs to all wh-questions in comparison to all yes/no-questions at the embedded object position. Visual inspection showed a sustained anterior negativity in response to all wh-questions. ANOVAs run in the window of 300-1300ms revealed a significant or marginal main effect of condition in all three arrays [midline: $F(1, 19) = 3.28, p = .086$; parasagittal: $F(1, 19) = 5.83, p = .026$; temporal: $F(1, 19) = 4.81, p = .041$], as well as a marginal condition x anteriority interaction in the temporal array [$F(2, 38) = 3.75, p = .063$]. These results indicated that wh-questions were more negative than yes-no questions, especially in the front-central regions of scalp.

To summarize, just as at the wh-modifier position, an anterior negativity was found at the embedded object position of wh-questions. This shows that the negativity started at the wh-modifier position, and continued through the embedded clause. This effect of anterior negativity cannot be an effect of wh-scope calculation (i.e. in support of the wh-scope calculation hypothesis), as at this point in the sentence, subjects had
not seen any Q-particle that would disambiguate the scope of the wh-element they had already seen. On the other hand, an effect of this type is entirely consistent with the wh-Q expectation hypothesis, in so far as the introduction of a wh-word triggers the parser's expectation for a Q-particle.
専務がどんなパソコンを買ったかと経理の係長が聞き言いましたか。

専務が新しいパソコンを買ったかと経理の係長が聞き言いましたか。

Figure 4.12 ERPs (n=20) recorded from all electrodes relative to a 200 ms prestimulus baseline. Shown are the ERPs to the embedded object position (‘what.kind.of/new PC-ACC’ in Table 4.1) of all wh- vs. yes/no-questions.
4.4.2.3 Between Embedded and Matrix Verbs

Let us now investigate the end of the stimulus sentences (i.e., the embedded verb itself and the matrix clause). The gloss of the relevant experimental sentences with the point of comparison capitalized is given below in (4.8).

(4.8)
a. [Director-N what.kind.of PC-A BOUGHT-THAT] ACCOUNTING-OF MANAGER-N SAID-Q
   ‘What kind of computer did the accounting manager say the director bought?’
b. [Director-N what.kind.of PC-A BOUGHT-Q] ACCOUNTING-OF MANAGER-N ASKED-Q
   ‘Did the accounting manager ask what kind of computer the director bought?’
c. [Director-N what.kind.of PC-A BOUGHT-THAT] ACCOUNTING-OF MANAGER-N SAID-Q
   ‘Did the accounting manager say that the director bought a new computer?’

At this position, only the matrix clause wh condition (4.8a) is expected to show the RAN effect, as it is the only condition that still has an unresolved expectation for a Q-particle. The embedded clause wh condition (4.8b) has presumably already fulfilled this expectation at the embedded verb-Q position, and there should be no expectation at all for a Q-particle in the yes/no control condition (4.8c). Thus comparisons were made between matrix vs. embedded clause wh-questions and between matrix clause wh- vs. yes/no-questions in three latency windows that covered one word (‘bought-Q/that’), two words (‘bought-Q/that accounting-of’), and four words (‘bought-Q/that accounting-of manager-N asked/said-Q’) from the embedded verb position. We extended the latency window to the end of the sentence in order to look for slow
potential effects between a wh-word and its associated Q-particle that would lend further support to the wh-Q expectation hypothesis.

4.4.2.3.1 Embedded vs. Matrix Clause Wh-Questions

Figure 4.13 shows ERPs to embedded vs. matrix clause wh-questions at the embedded verb ‘bought-Q/that’ position. This is the same figure as Figure 4.2 previously discussed in the wh-scope hypothesis section. To repeat, a RAN effect was seen in response to matrix clause wh-questions. ANOVAs performed in the 300-600 ms window showed no significant or marginal differences, but ANOVAs in the 300-900 ms window revealed a marginal condition x anteriority x hemisphere interaction in the temporal array \( F(2, 38) = 2.88, p = .071 \) indicating that matrix clause wh-questions in comparison to embedded clause wh-questions were more negative, especially over the right anterior regions of scalp. This time the direction of the RAN effect coincided with our hypothesis, although the effect was statistically marginal.

To recap, the wh-scope calculation hypothesis predicted that we should see the RAN effect at verb-Q positions that disambiguate wh-scope, namely, at the embedded verb-Q position of embedded clause wh-questions, and at the matrix verb-Q position of matrix clause wh-questions. On the other hand, the wh-Q expectation hypothesis predicted that we should see a RAN effect at the embedded verb position of matrix clause wh-questions, as subjects’ expectation for a Q-particle has not yet been met in this condition. Thus the wh-Q expectation hypothesis was also supported over the wh-scope calculation hypothesis in this comparison. Although there was a visible N400-
like effect to embedded clause wh-questions at posterior electrodes, ANOVAs run in
the window of 350-450 ms revealed no significant or marginal differences.
Embedded Wh
専務がどんなパソコンを買ったか経理の係長が聞きましたか。
[Director-N what.kind.of PC-A bought-Q] accounting-of manager-N asked-Q

Matrix Wh
専務がどんなパソコンを買ったと経理の係長が言いましたか。
[Director-N what.kind.of PC-A bought-that] accounting-of manager-N said-Q

Figure 4.13 ERPs (n=20) recorded from all electrodes relative to a 100 ms prestimulus baseline. Shown are the ERPs to the embedded verb position (‘bought-Q/that’ in Table 4.1) of embedded vs. matrix wh-questions.
Figure 4.14 shows ERPs to embedded vs. matrix clause wh-questions at the embedded verb and matrix subject modifier (‘bought-Q/that accounting-of’) positions. A RAN effect was observed to matrix clause wh-questions. ANOVAs run in the 300-1300ms window revealed a significant condition x hemisphere x anteriority interaction in the temporal array [F(2, 38) = 3.55, p = .042], indicating that matrix clause wh-questions were more negative than embedded clause wh-questions at right frontal regions.

In addition, an N400-like effect was observed in response to the matrix subject modifier (‘accounting-of’) position of embedded clause wh-questions at the midline. However, ANOVAs run in the time window of 1000-1100ms (350-450ms poststimulus onset of ‘accounting-of’) showed a significant condition x hemisphere x anteriority interaction only in the temporal array [F(2, 38) = 4.92, p = .014]. This appeared to be due instead to the RAN elicited by matrix clause wh-questions, as polarity and distribution were both opposite of what would be expected of an N400, i.e., over right temporal electrodes, the negativity was (a) in response to matrix (instead of embedded) wh-questions and (b) larger anterior than over posterior electrodes.

A P600-like effect was also observed in response to the matrix subject modifier (‘accounting-of’) position of matrix clause wh-questions at the midline regions. However, ANOVAs run in the time window of 1150-1450ms (500-800ms region of ‘accounting-of’) showed only a marginal condition x hemisphere x
anteriority interaction in the temporal array \([F(2, 38) = 3.17, p = .056]\), which again appeared to be due to the RAN response to matrix clause wh-questions.
Figure 4.14 ERPs (n=20) recorded from all electrodes relative to a 200 ms prestimulus baseline. Shown are the ERPs to the embedded verb and matrix subject position (‘bought-Q/that accounting-of’ in Table 4.1) of embedded vs. matrix wh-questions.
Figure 4.15 shows ERPs to embedded vs. matrix clause wh-questions between embedded and matrix verbs, as in ‘bought-Q/that accounting-of manager-N asked/said-Q’. A sustained anterior negativity was observed. ANOVAs run in the 300-2600 ms window confirmed the observation. There was a significant main effect of condition in the parasagittal [F(1, 19) = 4.95, p = .038] and temporal [F(1, 19) = 5.41, p = .031] arrays, as well as a significant or marginal condition x anteriority interaction in the parasagittal [F(4, 76) = 3.72, p = .066] and temporal [F(2, 38) = 4.44, p = .047] arrays. These results indicated that matrix clause wh-questions were more negative than embedded clause wh-questions, especially over anterior regions of scalp.

To summarize thus far, matrix clause wh-questions in comparison to embedded clause wh-questions elicited anterior negativity in three latency windows covering one word, two words, and four words from the embedded verb position. When the anterior negativity was lateralized (for the one-word and two-word comparisons), it was always right-lateralized. The overall results are thus far in line with the wh-Q expectation hypothesis.
Embedded Wh

専務がどんなパソコンを買ったか経理の係長が聞きましたか。

Matrix Wh

専務がどんなパソコンを買ったと経理の係長が言いましたか。

Figure 4.15 ERPs (n=20) recorded from all electrodes relative to a 400 ms prestimulus baseline. Shown are the ERPs to the from embedded to matrix verb position (‘bought-Q/that accounting-of manager-NOM asked-Q’ in Table 4.1) of embedded vs. matrix wh-questions.
4.4.2.3.2 Matrix Clause Wh- vs. Yes/No-Questions

Switching comparisons, Figure 4.16 shows ERPs to matrix clause wh- vs. yes/no-questions at the embedded verb ‘bought-that’ position. Similar to the matrix clause wh vs. yes/no comparison, a RAN effect was observed in response to matrix clause wh-questions. ANOVAs performed in the 300-600 ms window showed no significant or marginal differences, while ANOVAs in the 300-900 ms window revealed a significant condition x anteriority interaction in the midline array [F(2, 38) = 3.74, p = .035], as well as a marginal condition x hemisphere interaction in the parasagittal [F(1, 19) = 3.36, p = .083] and temporal [F(1, 19) = 3.67, p = .071] arrays. These results indicate that matrix clause wh-questions were more negative over right anterior regions.

Although on visual inspection there appeared to be a small N400-like effect to matrix clause wh-questions in response to matrix clause yes/no-questions at posterior electrodes, ANOVAs run in the window of 350-450 ms revealed no significant or marginal differences.\(^{12}\)

\(^{12}\) Although on visual inspection there also appeared to be a discrepancy in the 0-100 ms region between the two conditions, ANOVAs run in the window of 0-100 ms revealed no significant or marginal differences.
Matrix Wh
専務がどんなパソコンを買ったと経理の係長が言いましたか。
[Director-N what.kind.of PC-A bought-that] accounting-of manager-N __ said-Q

Matrix Yes/No
専務が新しいパソコンを買ったと経理の係長が言いましたか。
[Director-N new PC-A bought-that] accounting-of manager-N __ said-Q

Figure 4.16 ERPs (n=20) recorded from all electrodes relative to a 100 ms prestimulus baseline. Shown are the ERPs to the embedded verb position (‘bought-that’ in Table 4.1) of matrix wh- vs. yes/no-questions.
Figure 4.17 shows ERPs to matrix clause wh- vs. yes/no-questions at the embedded verb and matrix subject modifier (‘bought-that accounting-of’) position. A RAN effect was observed in response to matrix clause wh-questions. ANOVAs run in the 300-1300 ms window revealed a marginal condition x anteriority interaction in the parasagittal array \[F(4, 76) = 2.80, p = .075\] and a marginal condition x hemisphere x anteriority interaction in the temporal array \[F(2, 38) = 2.50, p = .095\], indicating that matrix clause wh-questions were somewhat more negative at right frontal regions. Although an N400-like effect to the matrix subject modifier (‘accounting-of’) of matrix clause yes/no-questions was visible observed at midline electrodes, ANOVAs run in the window of 1000–1100 ms (350-450 ms poststimulus onset of ‘accounting-of’) did not support the observation. There was a marginal condition x hemisphere x anteriority interaction in the temporal array \[F(2, 38) = 3.08, p = .058\], but this seemed to be due to the RAN response to matrix wh-questions, as there was no significant or marginal difference in the midline array. In the temporal array, polarity and distribution were both opposite of what would be expected of an N400, i.e., the negativity was (a) in response to matrix wh-questions (instead of matrix yes/no-questions) and (b) larger anterior than over posterior electrodes.
Figure 4.17 ERPs (n=20) recorded from all electrodes relative to a 200 ms prestimulus baseline. Shown are the ERPs to the embedded verb and matrix subject position (‘bought-that accounting-of’ in Table 4.1) of matrix wh- vs. yes/no-questions.
Figure 4.18 shows ERPs to matrix clause wh- vs. yes/no-questions between embedded and matrix verbs, as in ‘bought-that accounting-of manager-N said-Q’. A sustained anterior negativity was observed in response to matrix clause wh-questions at prefrontal electrodes. However, ANOVAs run in the 300-2600 ms window revealed no significant or marginal differences.

To summarize thus far, matrix clause wh-questions in comparison to matrix clause yes/no-questions elicited a RAN effect for the two latency windows that covered one word and two words from the embedded verb position, although the statistical results were somewhat weaker than matrix vs. embedded wh-question comparisons. In addition, there was a visual trend of anterior negativity continuing to the sentence end in the four-word window, although it was not statistically supported. We suspect that the lack of statistically significant difference might be due to the N400-like effect at the sentence end of matrix clause yes/no-questions. As briefly discussed in Section 4.4.1.2 with regard to the wh-scope calculation hypothesis, we suspect that subjects saw an unexpected Q-particle at the end of a sentence that could otherwise be declarative (see Section 4.5 Discussion).

In conclusion, we have seen (R)AN effects to matrix clause wh-questions in comparison to both embedded clause wh-questions and matrix clause yes/no-questions. Together with the effect of anterior negativity elicited by all wh-questions in the embedded clause region (Figure 4.12), the (R)AN effects seen in response to matrix clause wh-questions at the embedded verb and in the main clause seem to consistently support the wh-Q dependency hypothesis.
専務がどんなパソコンを買ったと経理の係長が言いましたか。

専務が新しいパソコンを買ったと経理の係長が言いましたか。

Figure 4.18 ERPs (n=20) recorded from all electrodes relative to a 400 ms prestimulus baseline. Shown are the ERPs to the region from embedded to matrix verb position (‗bought-that accounting-of manager-NOM asked-Q‘ in Table 4.1) of matrix wh- vs. yes/no-questions.
4.5 Discussion

4.5.1 RAN

To summarize, we have seen effects of anterior negativity in response to \textit{wh}-questions in multiple comparisons. The effects found in the present experiment (and the mono-clausal experiment discussed in Chapter 3) were right-lateralized in the majority of cases, although often statistically marginal. While they were sometimes bilateral in distribution, they were never left-lateralized. In the following subsections, we discuss what this might imply.

4.5.1.1 Wh-Scope Calculation or Wh-Q Dependency?

The results of the present experiment seem to support the wh-Q dependency hypothesis over the wh-scope calculation hypothesis. First, there was no local RAN at the sentence positions where wh-scope would be disambiguated. At the embedded verb position, although with marginal significance, matrix clause wh-questions in comparison to embedded clause wh-questions elicited a RAN effect. This coheres with what the wh-Q dependency hypothesis would predict (as a part of the continuous effect between wh and Q), while it is the exact opposite of what the wh-scope calculation hypothesis would predict (local scope calculation at the embedded verb-Q position). At the matrix verb position, there was a larger negative peak to matrix clause wh-questions in comparison to embedded clause wh-questions. However, the effect was different from the RAN effects seen in other comparisons in terms of latency (i.e., only in the 300-400ms latency window) and distribution (significant only in the midline array) (see Section 4.5.2 for more details on this effect). Secondly, both
types of wh-questions elicited RAN effects from the wh-modifier position to the corresponding Q-particle position (as shown in Figures 4.9-4.12 for all wh-questions in the embedded clause region and in Figures 4.13-4.18 for matrix wh-questions), as predicted by the wh-Q dependency hypothesis.

Therefore, the question raised at the end of Chapter 3 seems to have been answered, the wh-Q dependency (continuous expectation between wh and Q) hypothesis has been supported over the wh-scope calculation (local scope calculation at verb-Q) hypothesis. However, two differences between the results of the mono-clausal and bi-clausal experiments are worth mentioning. First, the present bi-clausal experiment revealed an anterior negativity at the wh-modifier position, while there was no anterior negativity seen at the wh-pronoun position in the mono-clausal experiment. Instead, only lexical effects influencing P200 amplitude and N350 amplitude and latency were found. However, this may have been because word length was not controlled for between wh-pronouns and demonstratives in the mono-clausal experiment. Recall that the wh-pronouns used in the mono-clausal experiment were shorter than the demonstratives, and that there was a later N350 peak to the longer demonstratives. This N350 peak to the demonstratives could conceivably have masked the anterior negativity to the wh-words.

Secondly, in the present experiment the negativity to wh-questions was longer lasting, while in the mono-clausal experiment the negativity was found only around sentence end. This cannot be attributed to differences in the distance between a wh-element and its corresponding Q-particle across the two experiments, as the average
number of words between wh and Q in the present study was 2.4, while that in the mono-clausal study was 3. But it may have something to do with the bi-clausal structure of the stimulus materials in the present study. In the present study, 73% of wh-questions were biclausal (100 out of 137 wh-questions); there was only one filler condition with mono-clausal wh-questions. Thus when subjects saw wh-elements, they may have immediately expected a biclausal structure, and that might have triggered the RAN effect earlier, because they might have anticipated that some (for matrix wh-questions) of the wh-Q dependencies were going to be long.

Although we have abandoned the possibility of RAN as an index of local wh-scope calculation at the verb-Q position, it has to be noted that we have not necessarily abandoned the possibility that there is a processing cost to wh-scope calculation. The (local) wh-scope calculation hypothesis says that wh-scope calculation happens at the Q-particle position, but this may not necessarily be the case. We could argue that the scope of wh-in-situ is calculated by the linkage with its corresponding Q-particle. Then the scope calculation is a long-lasting process, rather than something that happens locally at a verb-Q position. In this way, we could collapse the effects of wh-Q expectation and wh-scope calculation.

4.5.1.2 Other Possibilities: Prosody Effects Revisited

Although there was concern about a sentence-end wrap-up effect for the mono-clausal experiment, this was clearly not the case for the present experiment. There were (R)AN effects not only at sentence end, but also in response to the embedded
clause. The effects started at the wh-modifier position, five words before the final word of the sentence.

Aside from the sentence-end wrap-up effect, recall that we considered the possibility of a tacit prosody difference effect in Chapter 3. We suspected that subjects might have been reading stimulus sentences to themselves, and that the purported silent intonation difference across conditions might have been causing the RAN effect. In order to explore this possibility once again, a prosody analysis of sample tokens from the present study was also conducted.

Two native speakers of Japanese (one male and one female, Tokyo dialect) recorded 24 experimental sentences each (6 sets x 4 conditions) that were pseudo-randomized with 24 filler items. The digitized speech signals (44.1 kHz/16 bit sampling rate) were analyzed in terms of F0 contours. The F0 maxima of the embedded verb (the last word of the embedded clause) and the modifier of the matrix subject (the first word of the matrix clause) were measured and compared using repeated measures ANOVAs (treating item as a random factor) with Tukey HSD posthoc comparisons and paired one-tail t-tests.

Table 4.3 shows the F0 maxima of the relevant words, and Figure 4.19 shows typical F0 contours of the four experimental conditions. Recall from Chapter 3 that a wh-word in a wh-question is always accented and causes pitch reduction of the following words within a certain prosodic domain called the intonational phrase (Beckman & Pierrehumbert, 1986; Maekawa, 1991a, b; Beckman, 1996). Recall too

---

13 These were the same items as those used in the acceptability rating task in this chapter and Chapter 5 and the reading time task in Chapter 5.
14 Called “intermediate phrase” in Beckman & Pierrehumbert (1986) and Maekawa (1991a, b).
that unlike the mono-clausal experiment discussed in Chapter 3, wh-words used in the present experiment were noun-modifiers instead of entire noun phrases, such as ‘what.kind.of’ and ‘what-cup-of (how many cups of)’. Some of the non-wh-modifiers are lexically accented, while others are not.

Table 4.3  Mean F0 maxima (in Hz) of relevant words

<table>
<thead>
<tr>
<th>Condition</th>
<th>Male Embedded Verb</th>
<th>Male Matrix Subject Modifier</th>
<th>Female Embedded Verb</th>
<th>Female Matrix Subject Modifier</th>
</tr>
</thead>
<tbody>
<tr>
<td>Embedded Wh</td>
<td>92.8</td>
<td>129.2</td>
<td>222.5</td>
<td>287.8</td>
</tr>
<tr>
<td>Matrix Wh</td>
<td>98.0</td>
<td>122.8</td>
<td>219.2</td>
<td>230.0</td>
</tr>
<tr>
<td>Embedded Yes/No</td>
<td>97.5</td>
<td>130.7</td>
<td>252.7</td>
<td>298.5</td>
</tr>
<tr>
<td>Matrix Yes/No</td>
<td>100.0</td>
<td>127.5</td>
<td>234.8</td>
<td>283.3</td>
</tr>
</tbody>
</table>

Let us first consider a set of conditions in which a non-wh-modifier is not lexically accented, as in the case of *san-nen-no* ‘3-year-of’ (Figure 4.19c, d). In this case, there was no observable pitch reduction induced by the non-wh-modifier. On the other hand, there was a salient peak in pitch on the wh-modifiers such as *nan-nen-no* ‘what-year-of’ (Figure 4.19a, b), as well as a subsequent pitch reduction to the wh-conditions in comparison to the non-wh-conditions in the embedded clause region.

When a non-wh-modifier was lexically accented, such as with *ni-hai-no* ‘2-cup-of’ (Figure 4.19g, h), the lexical accent on the non-wh-modifier often created a similarly salient peak on the wh-counterpart (Figures 4.19e, f). Even in this case, the wh-conditions often showed a recognizable pitch reduction compared to the non-wh-conditions at the embedded verb position.
Figure 4.19  The F0 (fundamental frequency) contours of embedded clause wh-questions (a and e), matrix clause wh-questions (b and f), embedded clause yes/no-questions with unaccented (c) and accented (g) modifiers, and matrix clause yes/no-questions with unaccented (d) and accented (h) modifiers, as uttered by a female Tokyo Japanese speaker.
However, when ANOVAs were run on the \(F_0\) maxima at the embedded verb position (the last word of the embedded clause), the observation of pitch reduction in wh-conditions relative to non-wh-conditions in the embedded clause showed statistical significance only for the female speaker [male speaker: \(F(3, 5) = 1.12, p = .373;\) female speaker: \(F(3,5) = 11.25, p < .001,\) Tukey HSD: embedded wh, matrix wh < embedded yes/no]. At the matrix subject modifier position (the first word of the matrix clause), matrix clause wh-questions showed the lowest \(F_0\) among the four experimental conditions for both speakers. This, however, was again statistically significant only for the female speaker [male speaker: \(F(3, 5) = 1.35, p = .263;\) female speaker: \(F(1,5) = 25.70, p < .001,\) Tukey HSD: matrix wh < matrix yes/no, embedded wh, embedded yes/no].

In summary, consistent with the pitch reduction following a wh-word observed in the mono-clausal stimuli (see Section 3.5.2.3 in Chapter 3), there was a pattern of wh-induced pitch reduction in the scope of a wh-element in the present stimuli. At the embedded verb position, wh-conditions had a lower \(F_0\) compared to yes/no-conditions, and at the matrix subject modifier position (first word of the matrix clause), matrix clause wh-questions had the lowest \(F_0\) among all conditions. This pattern was generally observed for both male and female speakers, yet was only statistically significant for the female speaker, which might have been because females tend to have more elaborated pitch contours. It is especially interesting that matrix clause wh-questions showed the lowest \(F_0\) in the matrix clause region, at least visually for both speakers. This seems to suggest a correlation between wh-induced pitch reduction and
wh-scope, as the matrix clause region is still in the interrogative scope of a wh-element for matrix wh-questions (Cf. Deguchi & Kitagawa, 2002; Hirotani, 2003).15,16,17

15 Deguchi and Kitagawa (2002) point out that this correlation between wh-induced pitch reduction and wh-scope is a part of the general correlation between prosodic scope and focus scope. Maekawa (1991 a, b) also points out the general correlation between a pitch rise in a focused element and a subsequent pitch reduction in the following utterance.

16 Hirotani (2003) shows that when a wh-question like (i) has a prosodic boundary after the embedded Q-particle (indicated by a pitch reset between the embedded verb and the matrix verb (embedded verb F0 maximum < matrix verb F0 maximum)), there is a strong preference (84%) for an embedded clause wh-question response as in (ii). However, in the absence of such a prosodic boundary (no pitch reset), there is no preference between an embedded wh-question response (ii) (43%) and a matrix clause wh-question response (iii).

Our sample production data show that there was a pitch reset between the embedded verb and the matrix subject modifier for all conditions for the male speaker [embedded wh: t(5) = 8.25, p < .001; matrix wh: t(5) = 5.48, p = .001; embedded yes/no: t(5) = 5.09, p = .002; matrix yes/no: t(5) = 4.99, p = .002] and for all conditions except for matrix clause wh-questions for the female speaker [embedded wh: t(5) = 10.78, p < .001; matrix wh: t(5) = 1.27, p = .130; embedded yes/no: t(5) = 4.32, p = .004; matrix yes/no: t(5) = 5.01, p = .002]. This may be because the entire embedded clause is scrambled in our stimuli and prone to call for a pitch reset afterwards. From this, we can infer that embedded clause wh-questions in our stimuli were unlikely to have the prosodic cue for a matrix clause wh-question reading, indicating the intended embedded clause wh-question reading (Cf. Fodor (2002) for the argument for parser preference for the syntactic analysis associated with a default prosodic contour). In addition, as noted in footnote 1 in Chapter 3, it is not clear that much of the matrix wh-question response in Hirotani’s (2003) data was pragmatics-driven.

It is also interesting that in our stimuli, even matrix clause wh-questions showed a statistically significant pitch reset for the male speaker. This could again have been due to the scrambled embedded clause. Despite the pitch reset, however, the wh-induced pitch reduction seems to have continued at least visually for the male speaker. For the female speaker, there was no statistically supported pitch reset between the embedded and matrix clauses, and the pitch in the matrix clause region was statistically the lowest. This is consistent with the proposed correlation between wh-induced pitch reduction and wh-scope.

17 In terms of other wh-related tonal marking phenomena, Clements (1984) shows that in Kikuyu, when a wh-element is moved either partially or to its scope position, the verb in every clause between the moved wh-element and its gap loses a tonal downstep suffix. However, this phenomenon may be different from the wh-induced pitch reduction in Japanese discussed here, in that the downstep deletion in Kikuyu does not necessarily correspond with the interrogative scope of a wh-element (it stops at the intermediate clause if the wh-element is moved partially), and the downstep deletion Clements
Recall that there were continuing RAN effects in response to wh-questions. At the embedded clause region, all wh-questions elicited slow anterior potentials in comparison to all yes/no-questions. In addition, matrix clause wh-questions in comparison to embedded clause wh-questions and matrix clause yes/no-questions elicited RAN effects between the embedded and matrix verb positions. The domain of these RAN effects coincides with the domain of the wh-induced pitch reduction seen in our sample production data. Thus it is hard to deny the possibility that the RAN effect was induced by certain (tacit) prosodic properties. However, recall that the RAN effect was already seen at the wh-modifier position in the present experiment, and that there is a salient pitch peak at that position. Thus RAN is at least not directly related to pitch reduction itself. Nonetheless, we cannot deny the possibility that the prosodic pattern associated with wh-questions and their interrogative scope caused the RAN effect.

On the other hand, how could one ever tease apart the relative contributions of syntax, semantics, and prosody in sentences like this? Recall that in the previous section (Section 4.5.1.1), we concluded that the scope of wh-in-situ might be calculated by the linkage with its corresponding Q-particle. We could thus claim that subjects were imagining a lower pitch in the scope of a wh-element because they had not found the Q-particle they were looking for. Then the prosody effect, if it existed, might be subsumed under the effect of a subject’s searching for a Q-particle, or at least wh-scope. If wh-scope in Japanese is calculated by associating wh-words with Q-

---

(although we do not have any acoustic information on Kikuyu) seems to be the opposite of the pitch reduction discussed here.
particles (which seems perfectly plausible), and if pitch reduction reflects wh-scope, then we are just observing two different overt manifestations (brain response, pitch contours) of the same underlying cognitive phenomenon. In other words, there is no causal relationship that can be discerned here, as we could just argue that the brain response was driving the pitch reduction, since both (by hypothesis) indirectly reflect the establishment of wh-scope.

4.5.2 Other ERP Effects

Besides the (R)AN effect, we have seen several other effects in the present experiment. At the wh-modifier position, there were P200 and N400 effects for adjectives in comparison to donna ‘what.kind.of’ (see Figures 4.9 and 4.10). The P200 effect seems consistent with the finding discussed in Chapter 3, in that graphic complexity causes a larger P200. While donna ‘what.kind.of’ is all written in phonemic characters (kana) as どんな, 97.5% of the adjectives included Chinese characters (kanji) that are graphically more complex and take more strokes to write.

The N400 effect to open class adjectives compared to closed class ‘what.kind.of’ is also consistent with previous findings on the relationship between the N400 and lexical frequency. A corpus count of Asahi Shinbun (Amano & Kondo 2000), a popular Japanese newspaper, shows that ‘what.kind.of’ has a much higher frequency (27,715) than the mean frequency of the adjectives used (6,497). Thus the higher lexical frequency of ‘what.kind.of’ (Van Petten & Kutas, 1991) and its repeated presentation in the experiment (Van Petten, Kutas, Kluender, Mitchiner, & McIsaac,
would have lowered the N400 peak of the ERP to ‘what.kind.of’ and caused the N400 effect in response to adjectives.

From the above, it seems plausible to conclude that both effects are lexical. When graphic complexity and open vs. closed class differences in lexical frequency between wh- vs. non-wh-modifiers were more closely controlled in the numeral classifier pairs as in ‘what-cup-of’ and ‘2-cup-of’, both effects disappeared (Figure 4.11).

The N400-like effect in response to matrix yes/no-questions at the matrix verb-Q position (Figures 4.8 and 4.5) is unlikely to be due to the word length or frequency of ‘say’ and ‘ask’ verbs used for the two conditions, as each pair was matched for word length, and there was no statistically significant difference in the frequency of the two types of verbs (see Section 4.3.2). Instead, this may be an effect of seeing an unexpected Q-particle at the end of a sentence that otherwise could have ended as a declarative. If there is no Q-particle at the end of a matrix yes/no question, such as Senmu-ga atarashii pasokon-o katta-to keiri-no kakaricho-ga iimasita-ka ‘Did the accounting manager say that the director bought a new computer?’, the sentence would end as a declarative, as in Senmu-ga atarashii pasokon-o katta-to keiri-no kakaricho-ga iimashita. ‘The accounting manager said that the director bought a new computer.’ Although there were many more questions than declaratives in the experimental materials (275 sentences (79%) out of the 350 experimental materials), subjects might still have expected declaratives by default, in terms of their life experience. In particular, one filler condition had bi-clausal declaratives, as in ‘The
director of the museum said that the millionaire donated many pictures (see Table 4.2 (a)).’ Therefore, a Q-particle at the sentence-final matrix verb position was not very likely expected, and this may have caused the N400-like effect. The effect had a right central distribution in absolute values, a typical distribution for the N400.

At the sentence-final word (single word window), matrix clause wh-questions in comparison to embedded wh-questions showed a brief negative peak effect in the window of 300-400 ms, with a significant condition x anteriority interaction at the midline (Figure 4.4 in Section 4.4.1.2). From the distribution and latency, we have concluded that this is different from the RAN effect seen in other comparisons. This does not appear to be an N400 effect either, due to its anterior distribution and rather short latency. Unlike the Q-particle at the end of matrix yes/no-questions, subjects must have been satisfied to see the Q-particle they had been expecting at the end of matrix clause wh-questions. From this perspective as well, this effect seems unlikely to be a surprise-related N400 effect.

We speculate here that this may be a special effect of wh-scope integration across a clause boundary. We did not see this type of effect at the sentence-final verb-Q position of mono-clausal wh-questions, or at the embedded verb-Q position of embedded wh-questions. Thus we assume that this would not be an effect of wh-scope calculation in general. However, it may be feasible to think that there is an extra processing cost to integrating wh-scope across a clause boundary (see Kluender & Kutas, 1993b and Kluender, 1998 for a discussion of a processing cost for holding
and integrating a wh-filler across a clause boundary). Future research is needed to clarify this issue.

4.5.3 Are syntactically distinct languages processed the same way in the brain?

The results from the mono-clausal (Chapter 3) and bi-clausal (present chapter) experiments seem to suggest some similarities in the processing of wh-in-situ (Japanese) and wh-movement (English and German), in terms of both ERP components and general processing strategies. With respect to ERP components, we have seen that both filler-gap dependencies and wh-Q dependencies elicit anterior negativity, often in the form of slow potentials. With respect to processing strategies, we have seen that the parser seems to be expecting and actively seeking out a Q-particle following the appearance of a Japanese wh-word well before the Q-particle position, just like the parser actively tries to associate an English wh-element with its gap position (Active Filler Strategy: Frazier & Clifton, 1989).

4.5.3.1 Slow Anterior Negative Potentials

Slow anterior negative potentials have generally been associated with working memory load caused by both sentential and non-sentential linguistic stimuli. With respect to sentential stimuli, the ERP studies on filler-gap dependencies mentioned in Section 2.1.2 of Chapter 2 (King & Kutas, 1995; Müller, King, & Kutas, 1997; Kluender & Münte, 1998; Fiebach, Schlesewsky, & Friederici, 2001) reported slow anterior negative potentials. In addition, Münte, Schiltz, and Kutas (1998) manipulated the chronological order of events with before vs. after, as shown in (4.9).

(4.9) Before/After the scientist submitted the paper, the journal changed its policy.
Münte et al. reported left-lateralized slow anterior negative potentials in response to *before* sentences, whose amplitudes were correlated with the reading memory span of subjects. With respect to non-sentential stimuli, Ruchkin, Johnson, Canoue, and Ritter (1990) reported anterior negative slow potentials associated with the working memory retention of consonant strings with a greater variety of consonants within the same number of digits (6). Ruchkin, Johnson, Grafman, Canoune, and Ritter (1992) also reported left anterior negative slow potentials when subjects held longer pronounceable non-words in their short-term memory. This negativity was sensitive to memory load.18

Therefore, it seems justifiable to interpret the anterior negative slow potentials found in our study as an index of working memory load due to a dependency between a wh-element and its corresponding Q-particle in Japanese, as the parser would have to actively maintain a wh-element until its scope was licensed by the Q-particle. This may well be similar to the working memory load for associating a wh-filler with its gap in wh-movement languages. However, the question remains: why RAN instead of LAN?

First, it should be borne in mind that ours is not the first or the only study to report bilateral and/or right-lateralized slow negative potentials in response to syntactic dependencies. The slow negative potentials reported in King and Kutas

18 In terms of long term memory and non-linguistic stimuli, Rösler, Heil, and Glowalla (1993) investigated associative learning procedures (the fan effect) and reported robust left anterior negativities related to the active retrieval and evaluation of information. The size of this frontal negativity increased with the difficulty of the information retrieval.
were always bilateral. Additionally, Müller et al. (1997) ran an auditory version of King and Kutas’ (1995) subject vs. object relative study and reported right-lateralized anterior negativity to object relative clause sentences compared to subject relative clause sentences. Further, it has to be noted that although the anterior negativity effects found in the present study were often right-lateralized, these negativities were always found bilaterally and never solely on the right side of the scalp.

However, to account for the lateralization difference, it might be a good idea to review the underlying cognitive process involved in the processing of wh-movement and wh-in-situ. With respect to wh-movement languages such as English and German, it has been argued that storage of a wh-filler in working memory and its retrieval for filler-gap assignment causes a LAN effect. More specifically, Kluender and Kutas (1993 a,b) and King and Kutas (1995) relate the effect to the processing of a filler without a grammatical function (such as “subject” and “object”) or a thematic role (such as “agent” and “patient”) until the parser hits the verb-gap position and is able to find this missing information. Fiebach et al. (2001) take a more syntactic approach and claim that LAN is linked to “a syntactic working memory process that maintains the filler actively available until the gap is licensed and the syntactic chain between filler and gap can be established (p. 327)”. In essence, these researchers all claim that the process of holding a filler that is temporarily unlicensed (in terms of either grammatical function, thematic role, or syntactic chain) taxes working memory.
Unlike English or German, there is no extra ambiguity regarding the grammatical function or thematic role of wh-elements compared to their non-wh counterparts in Japanese. What is ambiguous about Japanese wh-elements is their interrogative scope until they are linked with their corresponding Q-particles, as discussed in Section 4.1, the Introduction. It might be the case that the process of holding a wh-element without knowing its scope also taxes working memory. For both wh-movement and wh-in-situ languages, the parser has to hold something with missing information for some extent of time, yet the type of missing information is different. What is unresolved about a wh-filler is fairly syntactic in nature, such as grammatical function, thematic role (although this is partially semantic), and structural chain formation. On the other hand, what is unresolved about a wh-in-situ is more logico-semantics in nature, namely, its interrogative scope. Although both kinds of information are needed for the successful interpretation of wh-questions, the computation of syntactic and logico-semantic properties may involve different neural generators. This may have contributed to the difference between left- and right-lateralization.

---

19 See Chapter 2 for the discussion on the attachment ambiguity of Japanese argument NPs in general.
20 One thing to consider is a direction the scope-licensor. As discussed in (3.7) in Chapter 3, in English a wh-element is displaced to the left of its gap and marks its scope, while in Japanese a Q-particle is placed to the right of the corresponding wh-element and marks its scope. Structurally, English and Japanese wh-words are placed in a phrase structure as shown in (i), up in the tree for English wh and down in the tree for Japanese wh.
4.5.3.2 Active Filler Strategy: Incremental Processing

Recall in Section 4.5.1.1 that we have concluded that the scope calculation is carried out by a long-distance linkage between a wh-element and its corresponding Q-particle. In other words, the parser seems to be expecting and actively seeking out a Q-particle following the appearance of a Japanese wh-word well before the Q-particle position, rather than waiting until the Q-particle position.

Aoshima, Phillips, and Weinberg (in press) also showed evidence for the incremental processing of Japanese wh-elements. Recall from Chapter 3 (3.12) that Aoshima et al. observed a TME slowdown even when a wh-element had to be interpreted as scrambled out of the embedded clause, as repeated here in (4.9).

This corresponds to the fact that the lower position is a thematic position while the higher position is a scope-licensing position. A Japanese wh-element receives a thematic role from its verb in the VP and then look for a scope-marker up in the tree to the right. On the other hand, an English wh-element receives its scope higher in the spec of CP and then look for a missing thematic role in the VP to the right.
(4.9)

a. Scrambled, Question Particle (Embedded clause wh-question)

\[
\text{EMBEDDED VERB}
\]
\[
\text{Dono-seito-ni tannin-wa [kocho-ga hon-o yonda-ka] toshoshitsu-de shisho-ni iimashita}
\]
which-student-D teacher-T principal-N book-A read-Q library-at librarian-D said

‘The teacher told (to) the librarian at the library to which student the principal read a book’

b. Scrambled, Declarative Complementizer (Matrix clause wh-question)

\[
\text{EMBEDDED VERB}
\]
\[
\text{Dono-seito-ni tannin-wa [kocho-ga hon-o yonda-to] toshoshitsu-de shisho-ni iimashita-ka.}
\]
which-student-D teacher-T principal-N book-A read-Q library-at librarian-D said-Q

‘To which student did the teacher tell (to) the librarian at the library that the principal read a book?’

They reported that the reading time at the embedded verb position was longer for matrix clause wh-questions (4.9b) than for embedded clause wh-questions (4.9a) and concluded that the parser prefers to interpret the scrambled wh-element as a part of the embedded clause and to associate the wh-element with the embedded verb-Q position, even though that would involve a clause external scrambling interpretation of the wh-element.

Aoshima et al. (in press) also showed that Japanese readers try to associate a fronted wh-phrase with the embedded clause. As seen in (4.10), they reported a slowdown at the embedded dative NP position for the condition with a scrambled
dative wh-NP in the matrix clause (4.10b), in comparison to the condition with no
dative wh-NP (4.10a).

(4.10)

a. Control

```
どの生徒が 担任に 校長が 図書室で 司書に 本を 読んだと 言いましたか。
Dono-seito-ga tannin-ni [kocho-ga toshoshitsu-de shisho-ni hon-o yonda-to] iimashita-ka
which-student-N teacher-T principal-N library-at librarian-at book-A read-that said-Q
```

‘Who said to the teacher that the principle read the book to the librarian at the
library?’

b. Scrambled

```
どの生徒に 担任は 校長が 図書室で 司書に 本を 読んだと 言いましたか。
Dono-seito-ni tannin-wa [kocho-ga toshoshitsu-de shisho-ni hon-o yonda-to] iimashita-ka
which-student-D teacher-T principal-N library-at librarian-at book-A read-that told-Q
```

‘To which student did the teacher say that the principle read the book to the
librarian at the library?’

They argued that this is a Japanese version of a filled-gap effect (Stowe, 1986), in that
the parser wanted to associate the filler with a possible gap position, but the position
had already been filled by something else, causing a slowdown. They argued that as
the imaginary gap-creation takes place before any verb position, sentence structures
are built incrementally before critical verbal heads.21

---

21 Aoshima et al. (in press) focused on associating the scrambled dative wh-NP with the gap position,
rather than with the Q-particle. They argued that the Active Filler Strategy (Frazier & Clifton, 1989)
cannot account for their data, as it would make the parser posit the gap as soon as possible, namely, in
the matrix clause before the beginning of the embedded clause, contra to their data suggesting that the
parser prefers to associate the wh-NP with the embedded clause. To account for their results, Aoshima
et al. proposed a principle based parsing, in that the interpretation of a wh-phrase is motivated by the
incremental satisfaction of grammatical constraints, such as thematic interpretation and scope licensing,
Therefore, the data by Aoshima et al. and the ERP and reading time evidence shown in Chapters 3 and 4 all seem to point to the conclusion that the parser actively and incrementally tries to associate a Japanese wh-element with its Q-particle, just like the parser actively tries to associate an English wh-element with its gap position (Active Filler Strategy: Frazier & Clifton, 1989). This seems to suggest that both English and Japanese are processed incrementally.

4.5.4 To what extent do neural processing data reflect syntactic structure?

In terms of the relationship between linguistic theory and brain responses, recall from Chapter 3 that linguistic theories argue that the scope of a wh-in-situ needs to be licensed by something in the Spec or head of its interrogative clausal CP. The Q-coindexing analyses focus on the relationship between wh-elements and Q-particles, in that wh-in-situ either requires a Q-particle in [C, CP] (Nishigauchi 1990; Cheng 1991) or needs to be bound by a Q-operator by spec-head agreement with the Q-particle in CP (Baker 1970; Pesetsky 1987; Watanabe 1992; Aoun and Li 1993; Cole and Hermon 1994) in order to have its scope licensed. If we assume rightward movement of wh-in-situ and rightward projection of Spec, as shown in Figure 4.20, and that the parser tries to associate the filler with the embedded clause, as the embedded verb appears earlier than the matrix verb.

It has to be noted, however, that the parser reportedly prefers to associate a non-wh dative NP with the matrix clause rather than the embedded clause (Kamide & Mitchell, 1999, although their matrix clause interpretation does not call for a scrambled structure like Aoshima et al.’s), and that the dative case marker -ni gives the parser much information about the thematic role of the NP. Thus the need for thematic interpretation by association with the gap position does not seem to be the major driving force in Aoshima et al.’s data. Instead, the need for wh-scope licensing by association with the Q-particle seems to be a more likely factor. We can shift our focus from the filler-gap association to the wh-Q association, and say that since the parser tries to associate the wh-filler with a Q-particle as soon as possible, it prefers to associate it with the embedded clause, which has the earliest possible host for a Q-particle (i.e., embedded verb). If this is the case, then the Active Filler Strategy is actually the driving force for the parser’s preference to associate the dative wh-NP with the embedded clause in Aoshima et al.’s data.
these claims basically entail a relationship between a wh-element and its Q-particle at the end of the clause. The LF-movement analyses (Huang 1982; Lasnik and Saito 1984; Nishigauchi 1990) can also be fit into this model, in that a wh-element moves to the Spec of CP and binds its trace at LF, again in spec-head agreement with a Q-particle in C.

\[
\text{CP} \quad \text{Spec} \\
\quad \text{C'} \\
\quad \text{IP} \\
\quad \ldots\text{wh}.. \\
\quad \text{C} \\
\quad \text{Q-particle}
\]

Figure 4. 20 Relationship between wh-in-situ and Spec, CP

This seems to be quite relevant to the (R)AN effects found between Japanese wh-elements and Q-particles in the present study. The parser actively expects a Q-particle after seeing a wh-element, and this is probably due to the need to determine its scope. We have abandoned the “RAN as an index of local wh-scope calculation at the verb-Q position” hypothesis, yet we could still say that the scope of a wh-element is licensed by the long-distance relationship between a wh-element and its corresponding Q-particle, as indexed by (R)AN.

In this sense, our result is compatible with both Q-coindexation and LF-movement analyses. This is plausible since (as pointed out in Chapter 3) in both analyses the scope of a wh-in-situ needs to be licensed by something in the CP that marks its scope (or COMP, to use a term that covers both Spec and C). Therefore, we can say that there is in general a dependency between wh-in-situ and COMP, and the
wh-Q dependency observed in processing studies mirrors in a more or less direct way the wh-COMP dependency posited by linguistic theory.

Therefore, the present study on the processing of bi-clausal Japanese wh-questions discussed in this chapter, in conjunction with the mono-clausal study discussed in Chapter 3, suggests a reliable neural processing correlate (as indexed by RAN) of the dependency between wh-elements and Q-particles. Theoretical claims about wh-in-situ argue for a mechanism inside COMP that licenses the scope of a corresponding wh-element, either by a Q-particle, a Q-operator, or the moved wh-element itself. This is consistent with our data showing that the brain actively seeks a Q-particle in COMP. From all of the above, we can say that the brain responses we measured reflect the complex relationship between a wh-in-situ and its licensing COMP as proposed in various linguistic theories.

4.6. Conclusion

In summary, we investigated how the brain responds to the processing of bi-clausal Japanese wh-questions. There was no clear local RAN effect at the embedded verb-Q position in embedded clause wh-questions or at the matrix verb-Q position in matrix clause wh-questions. This seems to suggest no clear effect of local wh-scope calculation. On the other hand, there were continuing (R)AN effects in response to wh-questions. At the embedded clause region, all wh-questions elicited slow anterior potentials in comparison to all yes/no-questions. Matrix clause wh-questions, in comparison to embedded clause wh-questions, elicited (R)AN effects between the embedded and matrix verb positions. Matrix clause wh-questions, in comparison to
their matrix clause yes/no-question counterparts, also elicited RAN effects at the embedded verb position, with a visible trend continuing through to sentence end.

This pattern of results was consistent only with the wh-Q dependency hypothesis as opposed to the wh-scope calculation hypothesis; there seems to be a reliable neural processing correlate of the dependency between wh-elements and Q-particles in Japanese. This effect is similar to the effect indexing the dependency between wh-elements and their gaps in wh-movement languages, but with a right-rather than left-lateralized distribution. Although we have abandoned the wh-scope calculation hypothesis in terms of a local scope calculation process at a verb-Q position, we are still left with the possibility that the scope of a wh-element is licensed by the incremental formation of a long-distance relationship between a wh-element and its corresponding Q-particle, which is indexed by RAN. Finally, RAN, as an index of the wh-Q dependency, can also be interpreted as the brain’s reflection of wh-COMP dependencies in linguistic theories.
Chapter 5: Reading Time Measures of Bi-Clausal Japanese Wh-Questions and Methodological Comparisons

5.1 Introduction

5.1.1 Miyamoto and Takahashi (2001, 2002b) (M&T)

This chapter examines the reading time processing of bi-clausal Japanese wh-questions and compares data collected using different methodologies. Recall that Miyamoto and Takahashi (2001, 2002b) (hereafter M&T) compared the reading times of embedded and matrix clause wh-questions with their non-wh counterparts as shown in (5.1).

(5.1) M&T’s Stimuli

a. Embedded QP/wh object (Embedded clause wh-question)

\[
\text{EMBEDDED VERB} \\
\text{専務がどんなパソコンを使っているか係長が聞いたの。} \\
[Senmu-ga donna-pasokon-o tsukatteiru-ka] kakaricho-ga __ kiita-no. \\
\text{director-N what.kind.of-computer-A is-using Q supervisor-N asked-Q} \\
\text{fast} \\
\text{‘Did the supervisor ask what kind of computer the director is using?’}
\]

b. Embedded affirmative complementizer/wh object (Matrix clause wh-question)

\[
\text{EMBEDDED VERB} \\
\text{専務がどんなパソコンを使っていて係長が言ったの。} \\
[Senmu-ga donna-pasokon-o tsukatteiru-to] kakaricho-ga __ itta-no. \\
\text{director-N what.kind.of-computer-A is-using that supervisor-N said-Q} \\
\text{slow} \\
\text{‘What kind of computer did the supervisor say the director is using?’}
\]
c. Embedded QP/NP object (NP-Q)

EMBEDDED VERB

専務が 新しいパソコンを 使っているか 係長が 聞いた。
Senmu-ga atarashii-pasokon-o tsukatteiru-ka kakaricho-ga kiita.

‘The supervisor asked whether the director is using the new computer’

d. Embedded affirmative complementizer/NP object (NP-‘that’)

EMBEDDED VERB

専務が 新しいパソコンを 使っていると 係長が 言ったの。
Senmu-ga atarashii-pasokon-o tsukatteiru-to kakaricho-ga itta.

‘The supervisor said that the director is using the new computer’

M&T compared matrix vs. embedded clause wh-questions at the embedded verb position and found that matrix clause wh-questions (5.1b) were read more slowly than embedded clause wh-questions (5.1a), while embedded QP/NP object sentences (5.1c) were read more slowly than affirmative complementizer/NP object sentences (5.1d).

M&T argued that a wh-word creates an expectation for interrogative typing of its clause by a Q[uestion]-particle (Cheng, 1991) while a non-wh-word creates an expectation for declarative typing of its clause by a non-Q-particle as the default. When this expectation is violated, it causes longer reading times. Sentences (5.1b) and (5.1c) were read more slowly because they violated this expectation. They called this delay in reading times the Typing Mismatch Effect (TME).
Besides the TMEs at the embedded verb position, M&T reported a shorter reading time on wh-objects than non-wh-objects, and a continuing TME until the sentence-end position for both wh- and non-wh- comparisons.

5.1.2 Aim of Present Study

Recall from Chapter 4 that we ran an ERP study on a modified version of M&T’s stimuli and found a right-anterior negativity in response to wh-questions. The aim of the study in this chapter was to run a self-paced reading time study on the modified stimuli and see if we could still replicate M&T’s data even with the modifications. In addition, we wanted to compare the obtained reading times to our ERP and acceptability rating data and see if there was any correlation between the data collected with different methodologies. In other words, we were interested in testing if a longer/shorter reading time would be related to a certain ERP effect, and if a lower acceptability rating would be related to longer reading times or a particular ERP effect.

5.1.3 Predictions

We predicted that we would on the whole replicate M&T even with our modifications and see TMEs as well as shorter reading times on wh-objects compared to non-wh-objects in our data. The following changes, however, were predicted due to the modifications.

First, M&T mostly used open-class adjectives as non-wh-modifiers of the embedded object. In the present study one half of the stimuli contained closed-class numeral classifier pairs, such as nan-bai-no ‘what-cup-of (how many cups of)’ and ni-hai-no ‘2-cup-of’, while the other half contained donna ‘what.kind.of’ vs. adjective
pairs, such as *donna* ‘what.kind.of’ and *atarashii* ‘new’. In this way, we were able to test whether M&T’s shorter reading times on wh-objects were due to an open-class vs. closed-class distinction, in that closed-class words may have been easier to process than open-class words, and thus may have taken less time to read (see Garnsey, 1985, for a larger N400 to open-class words than closed-class words that were matched for frequency and word length). If we were to see shorter reading times to wh-classifier objects than to non-wh-classifier objects, which are both closed-class items, we would be able to say that the difference is not due to an open-class vs. closed-class distinction.

Second, M&T presented embedded object NPs such as *donna pasokon-o* ‘what.kind.of PC-ACC’ as one unit. However, in the present study we separated such cases into two presentations as in *donna* ‘what.kind.of’ and *pasokon-o* ‘PC-ACC’, in order to increase the temporal distance between wh-words and the embedded or matrix verb-Q complex and be consistent with the overall style of word (bunsetsu) by word presentation. In addition, a modifier to the matrix subject, such as *keiri-no* ‘accounting-of’, was added to the matrix subject noun to extend the time before the matrix verb-Q complex. We predicted that this might increase the reading time on matrix clause wh-questions at sentence end, as there would be four regions between a wh-word and the verb-Q complex in the present study as opposed to two regions in M&T. We did not think this would affect the non-wh-comparison, as the mismatch effect at the embedded Q-particle position should not be increased as a function of the distance between a non-wh-object and a Q-particle. Subjects might be surprised to see
a Q-particle after seeing a non-wh-object at the embedded verb position, yet the surprise was predicted to disappear fairly soon.

Third, in the present study the non-wh conditions as well as the wh-conditions had interrogative endings so that we would have identical lexical items across all the four conditions, as opposed to the declarative endings used in M&T.\footnote{Other minor modifications in the present study included that the embedded verbs were all in the simple past tense and the matrix verbs were in the neutral polite form \textit{ii-mashi-ta-ka} `said-POL-PAST-Q', instead of the casual form \textit{i-hta-no} `say-PAST-Q', to better suit written presentation.} We predicted that this might increase the reading times on the non-wh-conditions at sentence end, as they could otherwise have ended as declaratives (recall from Chapter 4 that matrix clause yes/no-questions showed an N400-like effect at sentence end).

As for the comparisons among the data from the three (reading time, ERP, and acceptability rating) experimental methodologies, we predicted that they would be correlated in some fashion, along the lines of methodological correlations reported between reading time and ERP measures (King & Just, 1991; King & Kutas, 1995), ERP and rating measures (Kluender & Kutas, 1993b; Pechmann, Uszkoreit, Engelkamp, & Zerbst, 1994; Rösler, Pechmann, Streb, Röder, & Henninghausen 1998;), and reading time and rating measures (Mazuka, Itoh, & Kondo, 2002; Warren & Gibson, 2002). However, we also suspected that different methods might reveal different aspects of processing, such as lexical-level processing and sentence-level processing, and further, that there might be some relationship between patterns of results and types of methodologies based on an on-line (reading times, ERPs) vs. off-line (ratings) distinction or a neural (ERPs) vs. behavioral (reading times, ratings) distinction.
5.2 Norming

A paper-and-pencil rating study on the experimental items was conducted to norm the experimental materials, as reported in Section 4.2 of Chapter 4 and repeated below. Stimuli consisted of four conditions of bi-clausal questions, namely, (a) embedded clause wh-questions, (b) matrix clause wh-questions, (c) embedded clause yes/no-questions, and (d) matrix clause yes/no-questions, as shown in Table 5.1 (see also Appendix 2). The embedded clause was scrambled (displaced) to the beginning of the sentence (gap position indicated by underscoring), which is a common construction for a heavy object in Japanese (Yamashita & Chang, 2001; Yamashita, 2002).
Table 5.1 Sample stimuli

a. Embedded clause wh-questions

専務がどんなパソコンを買ったか経理の係長が聞きましたか。

‘Did the accounting manager ask what kind of computer the director bought?’

b. Matrix clause wh-questions

専務がどんなパソコンを買ったと経理の係長が言いましたか。

‘What kind of computer did the accounting manager say the director bought?’

c. Embedded clause yes/no-questions

専務が新しいパソコンを買ったか経理の係長が聞きましたか。

‘Did the accounting manager ask if the director bought a new computer?’

d. Matrix clause yes/no-questions

専務が新しいパソコンを買ったと経理の係長が言いましたか。

‘Did the accounting manager say the director bought a new computer?’
Filler items consisted of four conditions of declarative or interrogative sentences, namely, (a) bi-clausal declaratives, (b) mono-clausal ditransitive declaratives, (c) mono-clausal ditransitive yes/no-questions, and (d) mono-clausal ditransitive wh-questions as shown in Table 5.2. Declarative sentences were used to prevent subjects from always expecting interrogative endings. Similarly, mono-clausal sentences were used to prevent subjects from always expecting bi-clausal constructions. Ditransitive sentences were used to lead subjects to expect dative objects as well.
Table 5.2  Filler

a. Bi-clausal declaratives

美術館の館長が資産家が多数の絵画を寄附したと述べました。
bijutsukan-no kancho-ga shisanka-ga tasu-no kaiga-o kifushita-to nobe-mashi-ta.
museum-of director-N millionaire-N many-of picture-A donated-that state-POL-PAST

‘The director of the museum said that the millionaire donated many pictures’

b. Mono-clausal ditransitive declaratives

花屋の店員が往年の大女優に花束を届けました。
hanaya-no tenin-ga onen-no daijoyu-ni hanataba-o todoke-mashi-ta.
flower.shop-of clerk-N past.year-of big.actress-D bouquet-A deliver-POL-PAST

‘The clerk of the flower shop delivered a former star actress a bouquet’

c. Mono-clausal ditransitive yes/no questions

ベテランの補導員は制服の少女を保護者に渡しましたか。
beteran-no hodouin-wa seifuku-no shojo-o hogosha-ni watashi-mashi-ta-ka.
veteran-of school.counselor-T school.uniform-of girl-A guardian-D hand-POL-PAST-Q

‘Did the veteran school counselor hand a girl in her school uniform to her guardian?’

d. Mono-clausal ditransitive wh-questions

贅沢な社長夫人は何着のドレスを高級店に注文しましたか。
zeitakuna shachou.fujin-wa nan-chaku-no doresu-o kokyuten-ni chumonshi-mashi-ta-ka.
luxury president.wife-T what-piece-of dress-A boutique-D order-POL-PAST-Q

‘How many dresses did the luxury wife of the president place an order of to the boutique?’
Twenty-four sets (a subset of the 200 sets used in the ERP experiment discussed in Chapter 4) of experimental sentences were placed in a Latin square design to create four parallel lists such that no one subject saw more than one sentence from each set. Twenty-four filler items (a subset of the 150 sets used in the ERP experiment discussed in Chapter 4) were added to each list, and then each list was pseudorandomized. 20 native speakers of Japanese (same as the participants for the reading time study following the reading time session\(^2\)) were asked to rate each item on a scale from 1 (odd) to 5 (natural), with ‘natural’ meaning ‘sounds natural to a native speaker, appropriate in ordinary relaxed conversation’. However, as a version of the magnitude estimation task (Stevens, 1956) that allows a more continuous and open-ended rating system than a five-point scale, subjects were instructed to imagine drawing a line whose length indicated how acceptable a given sentence was, and then to choose one of the fixed points (on a 1-to-5 scale) on a scantron that corresponded most closely to the length of their imagined line (Cowart, 1997, see Chapter 1 Section 1.2.2.1).

\(^2\) Subjects were presented with different lists of experimental materials for the reading and rating studies.
Figure 5.1 Mean acceptability ratings on experimental materials. The scale that subjects used went from 1 (odd) to 5 (natural).

The results are shown in Figure 5.1. The mean acceptability ratings on the four experimental conditions ranged between 3.0 and 3.6 [embedded-wh: 3.6 (sd 1.29); matrix-wh: 3.0 (sd 1.23); embedded yes/no: 3.3 (sd 1.22); matrix yes/no: 3.3 (sd 1.31)]. Repeated measures ANOVAs were run with condition as a within-group factor, and either subject (F1) or item (F2) as a random factor. There was a statistically significant or marginal difference among the four conditions [omnibus ANOVA: F1(3, 19) = 6.30, p = .001; F2(3, 23) = 2.34, p = .075]. Tukey HSD posthoc comparisons revealed that embedded clause wh-questions were significantly more acceptable than matrix clause wh-questions for both subject and item analyses. Separate paired comparisons showed that embedded clause wh-questions were significantly more acceptable than matrix clause wh-questions for both subject and item analyses [F1(1,
19) = 13.48, p = .002; F_2(1, 23) = 6.71, p = .015] and that embedded clause wh-questions were significantly more acceptable than embedded clause yes/no-questions for the subject analysis [F_1(1, 19) = 11.85, p = .003] but not for the item analysis [F_2(1, 23) = 2.14, p = .15]. The pairwise comparisons between matrix clause wh and yes/no-questions [F_1(1, 19) = 2.66, p = .119; F_2(1, 23) = 2.65, p = .115] and between embedded and matrix clause yes/no-questions [F_1(1, 19) = .00, p = 1.000; F_2(1, 23) = .01, p = .919] revealed no significant or marginal differences.

5.3 Reading Time Study

This section reports the results of our reading time study in relation to Miyamoto and Takahashi’s (2001, 2002b) data.

5.3.1 Methods

5.3.1.1 Subjects

Twenty (13 female) monolingual speakers of Japanese between 21 and 29 years of age (mean: 24) who had been outside Japan for less than two years were included in the study. They were the same group of people as the participants for the norming study but different from those who participated in the ERP studies discussed Chapters 2-4. Subjects had normal or corrected-to-normal vision, were right-handed, and had no neurological or reading disorders. Subjects were reimbursed for their time. All of the subjects also participated in the norming experiments for the mono-clausal (see Chapters 2 and 3) and bi-clausal (see Chapter 4 and Section 5.2 of the present

---

3 The total number of subjects actually run was 30. However, five subjects showed a low comprehension score (below 83%), and five subjects saw only a part of the stimuli due to experimenter error. Thus their data were discarded.
chapter) wh-questions after the reading time experiment during the same testing session, with a short break between experiments.

5.3.1.2 Materials

Same as those used in the norming study in Section 5.2.

5.3.1.3 Procedure

The experiment was conducted on a MacOS 7 computer running PsyScope software (Cohen, MacWhinney, Flatt & Provost, 1993) and using a button-box for input. Subjects were timed in a word-by-word self-paced non-cumulative moving-window reading task (Just, Carpenter, & Woolley, 1982). Stimuli were presented on the computer screen in Japanese characters one bunsetsu at a time. A bunsetsu consists of one free morpheme (lexical word or pronoun) and the bound morpheme/s associated with it (particles modifying the noun/verb), and will be referred to as a “word” hereafter. All sentences, including the filler items, were presented on a single line. Stimulus segments initially appeared as a row of dashes, and participants pressed the leftmost button of the button-box to reveal each subsequent region of the sentence. Before beginning the experiment, subjects were given a practice set of eight sentences. The experiment took subjects approximately 15 minutes.

In order to maintain subjects’ attention, comprehension questions were inserted at the end of every sentence. (5.2) gives an example. A verb followed by two nouns was presented, and subjects had to decide which noun was the subject of the verb in
the preceding sentence. Subjects were asked to respond to the questions by pressing one of the two rightmost buttons of the button-box.

(5.2) Comprehension Question

a. Stimulus:

専務が どんな パソコンを 買ったか 経理の 係長が 聞きましたか。

[senmu-ga donna pasokon-o katta-ka] keiri-no kakaricho-ga_kikimashita-ka?

director-N what.kind.of PC-A bought-Q accounting-of manager-N asked.POL-Q

‘Did the accounting manager ask what kind of computer the director bought?’

b. Question:

聞く

kiku

1. 専務
senmu

director

2. 係長
kakaricho

manager

Auditory feedback (chime or buzzer) indicated whether the answer given was correct or not.

5.3.1.4 Data Analysis

Analyses were conducted on comprehension task response accuracy, reading times, and correlations among reading times, ERP amplitudes, and acceptability ratings. If the participant did not answer the comprehension task correctly, corresponding data points were eliminated from further analyses (6 % of trials). In addition, reading times were trimmed so that data points beyond 4 standard deviations from the relevant condition x region cell mean were discarded (3.9% of trials). The means and analyses presented below are based on the remaining trials.

---

4 This task was adopted from Nagata (1993), and was the same method used by Miyamoto and Takahashi (2001, 2002b).
For each sentence position, an omnibus ANOVA with repeated measures was conducted with experimental condition as a within-group factor, and either subject ($F_1$) or item ($F_2$) as a random factor. When the omnibus ANOVA for a certain sentence position was significant, posthoc comparisons were run using Tukey HSD posthoc comparisons with a family confidence coefficient of .95 ($\alpha = .05$ for the set of all pairwise comparisons). As for planned comparisons, separate ANOVAs were run comparing embedded vs. matrix clause wh-questions and embedded vs. matrix clause yes/no-questions at the embedded verb position and the matrix subject position, in order to test if the Typing Mismatch Effect (TME) reported by Miyamoto and Takahashi (2001, 2002b) would replicate. Additionally, reading times to wh-objects and non-wh objects for the adjective pairs (‘what.kind.of computer’ vs. ‘new computer’) and the numeral quantifier pairs (‘what-cup-of coffee’ and ‘2-cup-of coffee’) were compared at the embedded object modifier and noun positions. An alpha level of .05 was used for all statistical tests, with a p-value of .10 considered marginally significant. Following a modified Bonferroni procedure that allows up to [number of conditions -1] contrasts to maintain the same alpha level (Keppel, 1982), an alpha level of .05 was also used for planned pairwise comparisons.

5.3.4 Results and Preliminary Discussion

5.3.4.1 Comprehension Task Accuracy

The mean correct response rate to comprehension questions was 94% (range: 83-100% across subjects and 80-100% across items). The average correct response
percentage did not differ significantly across the four conditions [F₁(3, 19) = .46, p = .711; F₂(3, 23) = .398, p = .754].

5.3.4.2 Self-Paced Reading Times

Figure 5.2 and Table 5.3 summarize mean reading times per region for the four experimental conditions.
専務がどんな新しいパソコンを買ったかと経理の係長が聞き言いましたか

**Figure 5.2** Mean reading times per region for the four experimental conditions.

**Table 5.3** Mean reading times per region for the four experimental conditions (SD).

<table>
<thead>
<tr>
<th></th>
<th>1 director</th>
<th>2 what/new</th>
<th>3 PC</th>
<th>4 bought-Q/that</th>
<th>5 account-ing</th>
<th>6 manager</th>
<th>7 said/asked-Q</th>
<th>Grand Mean (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Embedded Wh</td>
<td>705 (327)</td>
<td>537 (152)</td>
<td>594 (271)</td>
<td>636 (281)</td>
<td>604 (196)</td>
<td>576 (213)</td>
<td>630 (203)</td>
<td>612 (346)</td>
</tr>
<tr>
<td>Matrix Wh</td>
<td>736 (403)</td>
<td>574 (188)</td>
<td>637 (294)</td>
<td>650 (286)</td>
<td>699 (345)</td>
<td>636 (270)</td>
<td>979 (659)</td>
<td>702 (345)</td>
</tr>
<tr>
<td>Embedded Yes/No</td>
<td>704 (347)</td>
<td>649 (263)</td>
<td>668 (372)</td>
<td>711 (396)</td>
<td>671 (332)</td>
<td>590 (204)</td>
<td>743 (345)</td>
<td>677 (345)</td>
</tr>
<tr>
<td>Matrix Yes/No</td>
<td>672 (301)</td>
<td>681 (331)</td>
<td>663 (306)</td>
<td>659 (271)</td>
<td>607 (198)</td>
<td>606 (250)</td>
<td>808 (434)</td>
<td>671 (434)</td>
</tr>
<tr>
<td><strong>Grand Mean</strong></td>
<td>704 (346)</td>
<td>612 (251)</td>
<td>640 (313)</td>
<td>664 (312)</td>
<td>645 (279)</td>
<td>601 (236)</td>
<td>790 (459)</td>
<td>665 (459)</td>
</tr>
</tbody>
</table>
The four conditions did not differ significantly \(F_1(3, 19) = .96, p = .409; F_2(3, 23) = .59, p = .592\) in region 1 (the embedded subject region, ‘director-N’ in Table 5.1). This was expected as all the items were identical across conditions.

In region 2 (the embedded object modifier, ‘what.kind.of/new’), there was a significant effect of condition \(F_1(3, 19) = 9.67, p < .001; F_2(3, 23) = 7.94, p < .001\). Tukey HSD posthoc comparisons found a significant difference between the means of the following groups for both subject and item analyses: embedded wh, matrix wh < matrix yes/no; embedded wh < embedded yes/no. These differences indicated that wh-modifiers were read faster than non-wh-modifiers. Planned pairwise comparisons also showed significantly shorter reading times on wh-modifiers for both adjectival (‘what.kind.of’ vs. ‘new’) \(F_1(1, 19) = 8.81, p = .003; F_2(1, 23) = 12.66, p = .001\) and numeral quantifier (‘what-cup-of’ vs. ‘2-cup-of’) \(F_1(1, 19) = 15.83, p < .001; F_2(1, 23) = 9.61, p = .002\) pairs. This will be attributed to a discourse anchoring effect of non-wh-modifiers in Section 5.3.5.

In region 3 (the embedded object noun, ‘PC-A’), the four conditions did not differ significantly \(F_1(3, 19) = 1.60, p = .189; F_2(3, 23) = 1.57, p = .197\). Planned pairwise comparisons revealed a significant difference only in the subject analysis of the wh- vs. non-wh-numeral quantifier pair \(F_1(1, 19) = 4.48, p = .036; F_2(1, 23) = 1.93, p = .166\], indicating that wh-modifiers were again read faster than non-wh-modifiers likely due to a continuation effect from region 2. There was no significant or marginal difference for the adjectival pair \(F_1(1, 19) = .10, p = .749; F_2(1, 23) = 1.61, p = .206\).
In region 4 (the embedded verb, ‘bought-Q/that’), the four conditions did not differ significantly \[F_1(3, 19) = 1.47, p = .221; F_2(3, 23) = 1.08, p = .356\]. Planned pairwise comparisons did not show any significant or marginal differences either \[\text{embedded wh vs. matrix wh: } F_1(1, 19) = .22, p = .640; F_2(1, 23) = .51, p = .477; \]
\[\text{embedded yes/no vs. matrix yes/no: } F_1(1, 19) = 1.47, p = .227; F_2(1, 23) = 1.23, p = .270\].

In region 5 (the matrix subject modifier, ‘accounting’), there was a significant effect of condition \[F_1(3, 19) = 3.77, p = .011; F_2(3, 23) = 3.12, p = .026\]. Tukey HSD posthoc comparisons found significant differences among the means of the following groups for subject analyses: embedded wh, matrix yes/no < matrix wh, while no group means were significantly different for item analyses. Planned pairwise comparisons between embedded and matrix clause wh-questions showed significant differences \[F_1(1, 19) = 7.96, p = .005; F_2(1, 23) = 6.87, p = .01\], and planned pairwise comparisons between embedded and matrix clause yes/no-questions \[F_1(1, 19) = 3.22, p = .074; F_2(1, 23) = 3.10, p = .080\] showed marginal differences. These differences indicated that matrix clause wh-questions were read more slowly than embedded clause wh-questions, and embedded clause yes/no-questions were read marginally more slowly than matrix clause yes-no-questions. This supports Typing Mismatch Effects (TMEs) for both wh- and non-wh-comparisons, in that wh-‘that’ and NP-Q take more time to read.

In region 6 (the matrix subject position, ‘manager-N’), the four conditions did not differ significantly \[F_1(3, 19) = 1.67, p = .172; F_2(3, 23) = 1.22, p = .303\].
However, planned pairwise comparisons between embedded and matrix clause wh-questions showed a marginal difference by subjects \( [F_1(1, 19) = 3.82, p = .052] \) and a significant difference by items \( [F_2(1, 23) = 6.87, p = .010] \), indicating that matrix clause wh-questions were read more slowly than embedded clause wh-questions and that the TME to matrix clause wh-questions continued from the previous word. There was no significant or marginal difference between embedded and matrix clause yes/no-questions \( [F_1(1, 19) = .35, p = .557; F_2(1, 23) = .33, p = .567] \).

In region 7 (the matrix verb position, ‘asked/said-Q’), there was a significant effect of condition \( [F_1(3, 19) = 12.40, p < .001; F_2(3, 23) = 12.00, p < .001] \). Tukey HSD posthoc comparisons found statistically significant differences among the means of the following groups for both subject-wise and item-wise analyses: embedded wh, embedded yes/no, matrix yes/no < matrix wh; embedded wh < matrix yes/no. These differences indicated that matrix clause wh-questions took the longest time, while embedded clause wh-questions took the shortest time to read.

5.3.4.3 Comparison to Miyamoto & Takahashi (2001, 2002b) (M&T)

Figures 5.3 and 5.4 show the mean reading times per region for the present study (hereafter abbreviated as U) and Miyamoto and Takahashi’s (2001, 2002b) study (hereafter abbreviated as M&T)\(^5\). We compare and discuss the results of both studies in this section.

\(^5\) We thank Edson Miyamoto for providing us with M&T’s data.
Figure 5.3 Mean reading times per region for the four experimental conditions (present study).

Figure 5.4 Mean reading times per region for the four experimental conditions (Miyamoto & Takahashi, 2001; 2002b).
In the embedded subject region (‘director-N’, region 1 for U and M&T), experimental conditions did not differ significantly among each other in either study. This is only natural as all the words were identical across conditions.

In the embedded object NP region (‘what/new PC-A’, regions 2-3 for U and region 2 for M&T), wh-NPs were read faster than non-wh-NPs in both studies. This will be attributed to a discourse anchoring effect of non-wh-modifiers in Section 5.3.5.

In the embedded verb region (‘bought/is.using-Q/that’, region 4 for U and region 3 for M&T), there was no significant or marginal difference in U. On the other hand, in M&T’s data, matrix clause wh-questions were read more slowly than embedded clause wh-questions and embedded clause NPs (NP-Q) were read more slowly than matrix clause NPs (NP-‘that’), showing Typing Mismatch Effects (TMEs), although both differences were statistically marginal.

In the matrix subject region (‘(accounting) manager-N’, regions 5-6 for U and region 4 for M&T), the TME (longer reading times to matrix wh than embedded wh and longer reading times to embedded yes/no than matrix yes/no) appeared one word later in U than in M&T. In U’s Region 5, matrix clause wh-questions (wh-‘that’) were read significantly more slowly than embedded clause wh-questions (wh-Q), and embedded clause yes/no-questions (NP-Q) were read marginally more slowly than matrix clause yes/no-questions (NP-‘that’), indicating TMEs in both cases. The TME continued for both wh and non-wh comparisons in M&T’s region 4; the marginal effects for the previous region became significant in this region, apparently because of the one-word delay. The TME continued for wh-questions (longer reading times to
matrix wh (wh-'that’) than embedded wh (wh-Q)) but ended for yes/no-questions in U’s region 6, as the reading times to both types of yes/no-questions were essentially the same. This will be attributed to the different processing demands of wh- and yes/no-questions, in terms of continuing expectation for a Q-particle, in Section 5.3.5.

In the matrix verb region (‘said/asked(-Q)’, region 7 for U and region 5 for M&T), matrix clause wh-questions (wh-'that’) were read more slowly in both studies, although much more so for U. In Section 5.3.5, this will be attributed to the fact that there were two more presentations between wh-word and Q-particle in the present study compared to M&T. As for the non-wh-conditions, it has to be noted that M&T’s non-wh-conditions were declarative while U’s were interrogative. The interrogative endings of U seemed to have taken much longer to read than the declarative endings of M&T, probably because interpreting interrogatives that could have ended as declaratives yielded an extra processing cost (see Section 5.3.5).

In terms of the regions in between the embedded and matrix verbs (regions 4-7 for U and regions 3-5 for M&T), both studies showed a continuing TME for the wh-comparison, in that matrix clause wh-questions (wh-'that’) took longer to read than embedded clause wh-questions (wh-Q), with the longer reading time at sentence-end much more pronounced for U than for M&T (again, possibly because there were two more presentations between wh-word and Q-particle for the matrix clause wh-condition of the present study, see Section 5.3.5). As for the non-wh-comparison, a TME (longer reading times to NP-Q than NP-'that’) continued until the sentence-end (two words after the embedded verb position) in M&T, while the effect ended within
two words after the embedded verb position in U. This again may be due to the nature of the difference in processing wh- and yes/no-questions, in terms of continuing expectation for a Q-particle (see Section 5.3.5).

5.3.5 Discussion: The Present Study and Miyamoto and Takahashi (2001, 2002b) (M&T)

Overall, the present study replicated Miyamoto and Takahashi (2001, 2002b) (hereafter M&T) with some predicted differences.

First, non-wh-objects took a longer time to read than wh-objects in both studies. The present study shows that this slowdown was unlikely to be due to an open-class vs. closed class distinction. Although the numeral quantifier pairs such as ‘what-cup-of coffee’ and ‘2-cup-of coffee’ were both closed class items, non-wh-objects such as ‘2-cup-of coffee’ were still read more slowly. One possible account for a delay in processing non-wh-NPs is their need for establishing discourse reference. Establishing discourse reference in the current text (or “discourse anchoring”) is said to yield an extra processing cost (cf. Garrod & Sanford, 1994). Wh-NPs such as ‘what kind of computer’ and ‘how many cups of coffee’ do not need a discourse referent representation, while non-wh-NPs such as ‘new computer’ and ‘two cups of coffee’ do, and this may have delayed reading times. Alternatively, Miyamoto and Takahashi (2003) argued that the delay may have been due to the fact that adjectives in the non-wh-NPs required a contrast set in order to be interpreted. Miyamoto and Takahashi cited Crain & Steedman (1985), who showed that readers prefer interpretations with fewer presuppositions. They argued that the existence of old computers has to be presupposed in order to interpret ‘new computer’ while that is not the case for ‘what
kind of computer’, and this may have slowed reading times in the former case. The same thing can be said for ‘how many cups of coffee’ vs. ‘two cups of coffee’ as well, in that ‘two cups of coffee’ may have to be contrasted with ‘three cups of coffee’ in order to be interpreted. In either case, non-wh-NPs may need some type of anchoring in relation to discourse entities (discourse reference, contrastive set) in order to be interpreted, and this may have incurred an extra processing cost.

Second, both studies showed TMEs. In the present study, the effects appeared one word later than in M&T. However, the TMEs M&T reported did not become significant until one word after the end of embedded clause, either. Further, a one-word delay with a TME was observed in Aoshima, Phillips, and Weinberg (in press) as well. The TME for the wh-comparison continued until the sentence-final position in the present and M&T’s studies, as predicted in the introduction. The TME for the non-wh-comparison also continued in M&T, but not in the present study, in which the effect was statistically marginal only at the matrix subject modifier position (the position immediately following the embedded verb. If we include the numerical (statistically unsupported) difference at the embedded verb position, the TME for the non-wh-comparison in the present study ended within two words after the embedded verb position. This may reveal the nature of the difference in processing wh- and yes/no-questions in terms of continuing expectations for a Q-particle. The parser would continuously maintain an unresolved expectation for a Q-particle until it was met for wh-questions, while the parser would recognize and react to a mismatch but soon recover from it for yes/no-questions.
At the sentence-final position, the present study showed overall longer reading times than M&T. In particular, matrix clause wh-questions took almost 1000 ms to read. This may have been because there were two more presentations between wh-word and Q-particle for the matrix clause wh-condition of the present study. One possible explanation is that the larger distance between the wh-phrase and the Q-particle incurred a larger integration cost\(^6\). Along with the ERP results, which suggested the possibility that the scope of a wh-element is licensed by the long-distance relationship between a wh-element and its corresponding Q-particle indexed by RAN (see Section 4.5.1 in Chapter 4), this may again suggest that wh-scope calculation is done via an incremental, long-distance linkage with a Q-particle. Both embedded and matrix clause yes/no-questions in the present study took longer to read than M&T’s NP-conditions, probably because interpreting interrogatives that could have ended as declaratives yielded an extra processing cost. 75% of the experimental materials were interrogative in the present study (50% stimuli and 25% filler), yet, as discussed in Section 4.5.2 in Chapter 4, subjects might still have imagined a declarative ending as the default, based on language input in their daily lives. This may have been another instance of a TME, in that there was an extra processing cost for an interrogative/declarative mismatch at the sentence final position of matrix clause yes/no-questions which could have ended as declaratives. The sentence-final

---

\(^6\) This is to some extent different from Gibson’s (1998, 2000) notion of an integration cost based on the number of intervening discourse referents (third-person NPs and verbs), as matrix clause wh-questions in the present study and M&T had an identical number of discourse referents between a wh-element and a Q-particle (‘PC-ACC’, ‘bought’, and ‘(accounting-of) manager-NOM’). What seems to be involved here instead is a temporal distance effect, in that how long ago subjects saw a wh-element is an important factor (E. Miyamoto, personal communication).
reading times to embedded clause wh-questions in the present study seemed to be roughly comparable to M&T’s embedded clause wh-questions, suggesting that the sentence length itself was not responsible for the longer reading times at the sentence-final position in the other three conditions in the present study.

5.4 Methodological Comparisons

This section compares the data from our reading time, ERP, and acceptability rating experiments. Table 5.4 and Figures 5.5-5.7 show the data used for methodological comparisons, namely, mean reading times across regions 1-7 and 4-7 and in region 7, mean ERP amplitudes for the four-word averages covering regions 4-7 collapsed across right anterior electrodes (Fp2, F4, F8), and acceptability rating scores. Note that subjects (20 native speakers of Japanese) who participated in the ERP experiment were a different group of people from those who participated in the reading time and rating studies (another group of 20 native speakers of Japanese).

<table>
<thead>
<tr>
<th></th>
<th>Mean Rating Scores</th>
<th>Mean Reading Times (ms)</th>
<th>Mean ERP Amplitudes (µV) (Fp2, F4, F8)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Regions 1-7</td>
<td>Regions 4-7</td>
</tr>
<tr>
<td>Embedded Wh</td>
<td>3.6</td>
<td>612</td>
<td>611</td>
</tr>
<tr>
<td>Matrix Wh</td>
<td>3.0</td>
<td>703</td>
<td>741</td>
</tr>
<tr>
<td>Embedded Y/N</td>
<td>3.3</td>
<td>677</td>
<td>679</td>
</tr>
<tr>
<td>Matrix Y/N</td>
<td>3.3</td>
<td>671</td>
<td>670</td>
</tr>
</tbody>
</table>
**Figure 5.5** Mean acceptability ratings on four experimental conditions.

**Figure 5.6** Mean reading times between embedded and matrix verbs (regions 4-7) of four experimental conditions.

**Figure 5.7** ERPs (n=20) to a right frontal (Fp2) electrode relative to a 400 ms prestimulus baseline after low-pass filtering (<1.5Hz.). Shown are ERPs between embedded and matrix verbs of four experimental conditions. Negativity is plotted up.
In order to see the statistical differences across conditions within the same methodology, omnibus ANOVAs with Tukey posthoc comparisons were run for each methodology (at the sentence final position for reading times and from 300-2600ms covering the sentence final four words for ERPs, which visually showed the largest difference across conditions).

As for ratings, there was a statistically significant or marginal difference among the four conditions [omnibus ANOVA: $F_1(3, 19) = 6.30, p = .001; F_2(3, 23) = 2.34, p = .075$]. Tukey HSD posthoc comparisons revealed that embedded clause wh-questions were significantly more acceptable than matrix clause wh-questions for both subject and item analyses.

As for reading times, there was a significant effect of condition [$F_1(3, 19) = 12.40, p < .001; F_2(3, 23) = 12.00, p < .001$]. Tukey HSD posthoc comparisons found statistically significant differences among the means of the following groups for both subject-wise and item-wise analyses: embedded wh, embedded yes/no, matrix yes/no < matrix wh; embedded wh < matrix yes/no. These differences indicated that matrix clause wh-questions took the longest time, while embedded clause wh-questions took the shortest time to read.

As for ERPs, there was a statistically significant or marginal difference among the four conditions [omnibus ANOVA: $F_1(3, 19) = 3.10, p = .034$]. Tukey HSD posthoc comparisons again revealed that matrix clause wh-questions were significantly more negative than embedded clause wh-questions.
Thus for all the three methodologies, it is consistent that matrix clause wh-questions call for the heaviest processing cost while embedded clause wh-questions call for the lightest processing cost. In the following subsections we correlate between different combinations of methodologies.

5.4.1 Reading Times and Rating Scores

Correlations between mean rating scores and mean reading times across regions 1-7, across regions 4-7, and in region 7 for the four experimental conditions were calculated, all of which showed significant correlations \([regions\ 1-7:\ r = -.966, p = .034;\ regions\ 4-7:\ r = -.998, p = .003;\ region\ 7:\ r = -.977, p = .023]\). This suggests a robust correlation between reading times and rating scores.

5.4.2 ERPs and Rating Scores

A correlation between mean rating scores and mean ERP amplitudes covering regions 4-7 (collapsed across right anterior electrodes, 300-2600ms window) for the four experimental conditions was calculated, which revealed a marginal \([r = .926, p = .073]\) correlation. This suggests some correlation between ERP and rating measures. Mean ERP for region 7 (in the latency window of 1950-2600ms of the four-word averages covering ‘bought-Q/that accounting-of manager-N asked/said-Q’) showed no significant or marginal correlation with rating scores.

5.4.3 Reading Times and ERPs

To start from the embedded clause region, recall that the reading times were shorter for wh-objects than for non-wh-objects for both adjective and classifier comparisons, possibly due to the discourse anchoring effect of non-wh-modifiers. On
the other hand, recall that the ERPs showed anterior negativity to wh-objects in comparison to non-wh-objects, again for both adjective and classifier comparisons (see Figures 4.9-4.12 in Chapter 4). Thus the results from reading time and ERP studies were, in this respect, the opposite of each other: wh-objects took less time to read but elicited anterior negativity. Although there was an N400 effect to non-wh-adjective modifiers in comparison to donna ‘what.kind.of’ (see Figure 4.10 in Chapter 4), this cannot be the sole reason for the longer reading times to non-wh-objects, as non-wh-objects took a longer time to read even in the classifier comparison in which the N400 effect was absent (see Figure 4.11 in Chapter 4). This will be attributed to the difference in underlying cognitive processes the two methodologies revealed in Section 5.4.4.

Unlike the above, the two methodologies coincide in the wh-comparisons around the embedded verb region. The reading times to matrix clause wh-questions were longer than those to embedded clause wh-questions and matrix clause yes/no-questions around the embedded verb region (‘bought accounting’), as shown in Figure 5.5. Recall that ERPs also showed right anterior negativity (RAN) to matrix clause wh-questions in comparison to embedded clause wh-questions or matrix clause yes/no-questions (see Figures 4.13-4.18 in Chapter 4) in this region, consistent with the reading time measures.

As for the non-wh-comparison around the embedded verb region, embedded clause yes/no-questions were read marginally more slowly than matrix clause yes/no-questions, which was considered to be a TME to a non-wh (NP-Q) condition. Since
this comparison was not considered in Chapter 4, ERPs in response to embedded and matrix clause yes/no-questions around the embedded verb region ‘bought-Q/that accounting-of’ were compared in order to look for a possible ERP correlate (see Figure 5.8). Although embedded clause yes/no-questions in comparison to matrix clause yes/no-questions visually showed anterior negativity, ANOVAs run in the 300-1300ms window revealed no significant or marginal difference.
Figure 5.8 ERPs (n=20) to all electrodes relative to a 200 ms prestimulus baseline. Shown are the ERPs to the embedded verb and matrix subject position (‘bought-that accounting-of’ in Table 5.1) of embedded vs. matrix clause yes/no-questions.
Moving onto the matrix clause region, in order to further investigate a possible correlation between reading time and ERP measures, we calculated the correlation between reading times across regions 4-7 and mean ERP amplitudes covering regions 4-7 collapsed across right anterior electrodes (from 300-2600ms) for the four experimental conditions. The correlation was barely marginal \( r = -.898, p = .102 \). However, this may have been due to the small number (four\(^7\)) of data points, as the correlation coefficient of -.898 is a fairly high number. We also calculated the correlation in region 7 and mean ERP amplitudes for region 7 (in the latency window of 1950-2600ms of the four-word averages covering ‘bought-Q/that accounting-of manager-N asked/said-Q’), which showed a marginal \( r = -.925, p = .074 \) correlation.

As discussed earlier, matrix clause wh-questions took the longest time to read, while embedded clause wh-questions took the shortest time to read (Figure 5.6), and this is consistent with the largest amplitude anterior negativity to matrix clause wh-questions and the smallest amplitude anterior negativity to embedded wh-questions (Figure 5.7). In addition, both embedded and matrix clause yes/no-questions took longer to read than M&T’s declaratives at the sentence-final position. This might have been because interpreting interrogatives that could have ended as declaratives might have exacted an extra processing cost. Recall that in Chapter 4 we saw an N400-like effect to matrix clause yes/no-questions in comparison to embedded clause yes/no-questions (Figure 4.8 in Chapter 4) at the sentence-final word. Thus there

\(^7\) We were unable to plot data points for each item or each subject individually, as our reading time and ERP experiments used different numbers of stimulus sentences (24 sets for reading times and 200 sets for ERPs) and two different groups of 20 subjects.
seems to be some ERP index of interpreting a non-wh sentence as interrogative, although then one might also expect to see an N400-like effect to embedded clause yes/no-questions at sentence end as well. In one case (matrix yes/no), there is no indication that the sentence is interrogative until sentence end. In the other case (embedded yes/no), there was already a TME effect at the embedded verb. Perhaps subjects subsequently revised their expectations in preference of an interrogative interpretation.

5.4.4 Discussion: Correlations among Reading Times, ERPs, and Ratings

Overall, the data from the three methodologies (reading times, ERPs, and rating scores) show a consistent pattern within the same methodology: matrix clause wh-questions call for the heaviest processing cost, while embedded clause wh-questions call for the lightest processing cost. We suppose this would be due to subjects’ unfulfilled expectation for a Q-particle as predicted by the wh-Q dependency hypothesis, which is demonstrated quite remarkably in all three methodologies.

When compared across methodologies, the data from three methodologies also correlate with each other fairly well. Acceptability ratings were significantly correlated with reading times across the entire sentence, between the embedded and matrix verbs, and on the matrix verb. The ratings were marginally correlated with ERP amplitudes at right anterior electrodes covering the regions between the embedded and matrix verbs. In addition, reading times in the sentence-final matrix verb region were marginally correlated with ERP amplitudes at right anterior electrodes covering the same region. The statistical results are not very robust, yet this
may be partly because we were unable to plot data points for each item or each subject individually, as our reading time and ERP experiments used different numbers of stimulus sentences (24 sets for reading times and 200 sets for ERPs) and two different groups of 20 subjects. Furthermore, although statistically non-significant, recall from Chapter 2 that scrambled wh- and non-wh- questions were rated lower than their in-situ counterparts, and from Chapter 3 that scrambled and in-situ wh-questions were rated lower than their non-wh counterparts. Thus there seems to be some general correlation between the different methodologies. This is in a sense expected, as all three methodologies are designed to measure certain aspects of sentence processing.

Interestingly, however, there were some differences between the reading time and ERP measures in the present study. First, although wh-objects were read faster than non-wh-objects, ERPs showed anterior negativity to wh-objects in comparison to non-wh-objects. This discrepancy might have been due to the difference in underlying cognitive processes the two methodologies revealed. Reading times might have revealed lexical semantics-level processing in anchoring a non-wh-word to a discourse, while ERPs might have revealed sentence-level processing, which seemed to be a part of the (R)AN effect to wh-questions. Then there may have been a delay in picking up the sentence-level wh-Q dependency effect in reading times. Assuming that the slow reading times at the non-wh object are related to discourse anchoring as discussed in Section 5.3.5, we would expect an “ERP discourse anchoring component” (N400?, or whatever that may be) to non-wh-objects. However, ERPs seemed to have revealed a different aspect of the process, namely, the expectation for a Q-particle on wh-objects,
which may not have been as salient at that point in reading time measures. Second, there was no apparent ERP correlate of the TME for the non-wh comparison around the embedded verb position. This discrepancy is much harder to account for. We do not have any good explanation for this other than to say that the TME in the present reading time study was actually statistically marginal and thus not very strong.

Comparing off-line (rating) vs. on-line (reading time, ERP) measures, the off-line rating measures do not necessarily show the statistically significant differences that on-line measures might show. Recall that in the mono-clausal experiments in Chapters 2 and 3, the rating scores were coherent with the ERP measures (lower ratings with the conditions that elicited RAN or AN effects than with the control conditions), but the differences among rating scores never reached a significance level. In the present study, matrix wh-questions were rated significantly lower than embedded wh-questions, but matrix clause yes/no-questions and matrix clause wh-questions differed only numerically from each other. Unlike this, reading times revealed statistically significant differences between matrix clause wh-questions and the other three conditions, and ERPs revealed statistically supported RAN effects between matrix clause wh-questions and embedded clause wh-questions/matrix clause yes/no-questions. In addition, both reading time and ERP measures can be used to pinpoint at which part of the sentence a certain effect occurred, while that was not the case for ratings.

In terms of comparing neural (ERP) vs. behavioral (rating, reading time) measures, ERP measures might be generally more useful in showing the nature of
processing, in that the multidimensionality of ERPs (e.g. latency, polarity, distribution) can provide a more fine-grained picture of processing than ratings and even reading times. We saw slower reading times in various cases that yielded different brain responses, such as the increasing reading times between embedded and matrix verbs of matrix wh-questions as reflected by RAN, and the sentence-final increase in matrix yes/no-questions as reflected by an N400-like effect. We can investigate various phenomena in language processing with ERPs in a way that is not possible with reading times alone. For instance, we saw reading time differences among the four experimental conditions in regions 4-7, yet this does not directly tell us what the differences might be due to. On the other hand, ERPs revealed a (R)AN effect to matrix clause wh-questions, which we interpreted was a working memory effect, and an N400-like effect to matrix clause yes/no-questions, which we interpreted was a surprise effect due to seeing an unexpected Q-particle.

Interestingly, however, we have encountered one case in which ERPs do not give us more information than reading times do. Non-wh-objects were read more slowly than wh-objects, which we suspected may have been due to their need for discourse anchoring. On the other hand, ERPs showed an anterior negativity effect in response to wh-objects, which was presumably part of the (R)AN effect to wh-questions, but no apparent effect to non-wh-objects in comparison to wh-objects indexing their need for discourse anchoring. We have suggested that this was because the reading times showed a lexical semantic-level effect while the ERPs showed a sentence-level effect. In addition, we could not find any apparent ERP correlate to the
TME for the non-wh comparison. Given these data, it might be best to use both methodologies with the same stimuli and compare their results.

5.5 Conclusion

In conclusion, the present study generally replicated Miyamoto and Takahashi’s (2001, 2002b) reading time study in that we saw Typing Mismatch Effects (TMEs) for both wh- and non-wh comparisons, and shorter reading times to wh-objects than to non-wh-objects. In addition to the above, the present study revealed several new findings. First, shorter reading times to wh-objects was not due to an open class vs. closed class distinction, but rather to some specific properties of wh-words, possibly related to discourse anchoring. Second, the TME is local for non-wh-sentences, in that it begins in the embedded verb region, but then ends within two words. On the other hand, the TME continues until the parser finds a Q-particle for wh-questions. Finally, the continuing TME and sentence-final lengthening of reading time to matrix clause wh-questions (along with the ERP evidence discussed in Section 4.5.1 in Chapter 4) seem to suggest that the scope of a wh-element is licensed by an incremental, long-distance linkage between a wh-element and its corresponding Q-particle.

In addition, the present study revealed that rating, reading time, and ERP measures show a consistent pattern of results with respect to continuing wh-Q dependencies. In addition, the data from these methodologies generally correlate with each other. However, it was also found that a particular methodology may be useful in
showing a particular aspect of processing, in that reading times may show lexical semantic-level processing and ERPs might show sentence-level processing.
Chapter 6: Conclusions

6.1 Chapter Summary

This thesis has examined the processing of Japanese mono-clausal and bi-clausal wh-questions mainly in comparison to English, using event-related brain potentials (ERPs), reading time measures, and acceptability ratings.

Chapter 2 investigated the relationship between a displaced (scrambled) wh-element and its gap. Mono-clausal scrambled wh-questions were compared against mono-clausal wh-in-situ questions. Despite the difference in basic word order between English and Japanese, the experiment revealed patterns of brainwave responses to Japanese scrambled wh-questions to be similar to those found in response to wh-movement in English and German, i.e., slow anterior negative potentials between filler and gap (and beyond the gap) and phasic LAN and P600 effects around the gap positions. This replication of results points to the existence of universal parsing operations for filler-gap dependencies. This pattern of results is less compatible with a head-driven parser model (e.g., Pritchett, 1992) and more compatible with an incremental parser model (e.g., Inoue & Fodor, 1995). Further, the results are compatible with several processing models pertaining to filler-gap dependencies.

Chapter 3 began exploring the relationship between a wh-element and its corresponding Q-particle -ka. Mono-clausal Japanese wh-questions and their structurally equivalent yes/no-question counterparts were compared. The
experiment revealed lexical effects within wh-words, namely, enhanced P200 and N350 components, with the P200 effect correlated with stroke count, and the N350 effect correlated with word length in terms of both amplitude and latency. But more crucially, both scrambled and in situ mono-clausal wh-questions elicited right-lateralized anterior negativity (RAN) around sentence end, relative to their yes/no-question counterparts. The question remained, however, whether this effect would suggest a long-distance dependency between a wh-element and its Q-particle (wh-Q dependency hypothesis) or a local scope calculation process at the verb-Q complex position (wh-scope calculation hypothesis).

In order to solve the remaining question from Chapter 3 whether the wh-Q dependency hypothesis or the wh-scope calculation hypothesis was right, Chapter 4 examined the ERP responses to bi-clausal Japanese wh-questions with matrix and embedded scopes and their structurally equivalent yes/no-question counterparts. At the embedded clause region, all wh-questions elicited sustained anterior negativity in comparison to all yes/no-questions. Matrix clause wh-questions, in comparison to embedded clause wh-questions, elicited (R)AN effects between the embedded and matrix verb positions. Matrix clause wh-questions, in comparison to matrix clause yes/no-questions, also elicited RAN effects at the embedded verb position, with a visible trend continuing through to sentence end.

This pattern of results is consistent only with the wh-Q dependency hypothesis as opposed to the wh-scope calculation hypothesis. There seemed to be a reliable neural processing correlate of the dependency between wh-elements and Q-particles in
Japanese, which is similar to the ERP effects seen between wh-fillers and gaps in English, but with a right- rather than left-lateralized distribution. Although the wh-scope calculation hypothesis, which indicates a local scope calculation process at a verb-Q position, does not seem plausible, we could still argue that the scope of a wh-element is licensed by a long-distance incremental linkage between a wh-element and its corresponding Q-particle, which is indexed by RAN. Further, the RAN effects found in Japanese wh-questions are consistent with theoretical claims that the scope of a wh-in-situ needs to be licensed by a scope marker (Q-particle, Q-operator, or the moved wh-element itself) in Spec and/or head of the CP (or COMP, to use a term that covers both Spec and head of CP). Thus the results may also suggest that RAN, as a manifestation of the wh-Q dependency, can also be interpreted as the brain’s reflection of wh-COMP dependencies discussed in linguistic theories.

Chapter 5 examined the reading time processing of bi-clausal Japanese wh-questions, in order to explore the possible correlations among the three methodologies (ERPs, reading times, acceptability ratings) used in this dissertation, as well as to test whether the experiment would replicate the results of Miyamoto and Takahashi’s reading time study (2001, 2002b), based on whose stimuli the biclausal ERP study in this dissertation was designed. The results were generally consistent with the data by Miyamoto and Takahashi and replicated the Typing Mismatch Effect (TME), in which a wh-word creates an expectation for the interrogative typing of its clause by a Q-particle, while a non-wh-word creates an expectation for the affirmative typing of its clause by an affirmative particle, and a mismatch in typing causes a slowdown. The
results further revealed a continuing TME in reading times for matrix clause wh-questions. This, along with the continuous (R)AN effects discussed in Chapter 4, suggests again that the scope of a wh-element is licensed by a long-distance linkage between a wh-element and its corresponding Q-particle formed in an incremental fashion. In addition, the present study revealed that rating, reading time, and ERP measures show a consistent pattern of results with respect to continuing wh-Q dependencies. In addition, the data from these methodologies generally correlate with each other, with ratings showing the most coarse measures, followed by reading times and then ERPs. However, there was one case in which ERPs seemed to have revealed sentence-level processing while reading times seemed to have revealed lexical semantic-level processing.

6.2 Answers to Research Questions

In sum, the experiments in Chapters 2 through 5 have provided the following answers to the research questions posited in Chapter 1, repeated here in (6.1).

(6.1) Research questions addressed in this dissertation:

a. To what extent are syntactically distinct languages processed the same way in the brain?

b. How is the scope of wh-in-situ calculated in real-time processing?

c. To what extent do neural processing data reflect linguistic theory?

d. What is the nature of the relationship (if any) among the data obtained from different methodologies?

We examine each of these questions below.
6.2.1 Question 1: To what extent are syntactically distinct languages processed the same way in the brain?

Overall, the results suggest numerous similarities in how syntactically distinct languages, such as Japanese and English, are processed in the brain. We have seen this in terms of both ERP components and general processing strategies.

6.2.1.1 Similarities in ERP Components: LAN/RAN and P600

As for ERP components, we have seen (a) slow potential anterior negativity between scrambled fillers and their gaps (and beyond) as well as phasic P600 and LAN effects around the gap position for the processing of Japanese scrambling (Chapter 2) and (b) RAN effects between wh-elements and their scope licensing Q-particles for the processing of mono-clausal and bi-clausal Japanese wh-questions (Chapters 3 and 4). These effects are all similar to the ERP effects reported for filler-gap dependencies in English and German, namely, sustained (L)AN effects between filler and gap (and also beyond the gap) (King & Kutas, 1995; Münte, King, & Kutas, 1997; Kluender & Münte, 1998; Fiebach, Schlesewsky, & Friederici, 2001) and phasic LAN (Kluender & Kutas, 1993a,b; King & Kutas, 2005) and P600 (Kaan, Harris, Gibson, & Holcomb, 2000; Phillips, Kazanina, Wong, & Ellis, 2001; Phillips, Kazanina, Garcia-Pedrosa, & Abada, 2003) effects around the gap. With regard to the processing of scrambling, the pattern of results is remarkably similar to that reported for filler-gap dependencies in English, regardless of the word order difference between the two languages. In the case of wh-Q dependencies, the anterior negativity was often right-lateralized, rather than left-lateralized, but one thing to note is that these negativities were always found bilaterally and never solely on the left or right side of
the scalp. Although the exact nature of the distributional difference will have to be scrutinized in the future (see Section 6.3.2), the RAN effects found in the present study are certainly some form of anterior negativity. Thus we can safely say that in relation to both its canonical position and its scope-marking Q-particle, a Japanese wh-element elicits anterior negativity, which is likely to indicate working memory load and a dependency between the wh-element and its gap or corresponding Q-particle.

6.2.1.2 Similarities in Processing Strategies: Incremental Processing

The present study strongly suggests that although Japanese is a verb-final language, it is nevertheless processed incrementally. We can see evidence for this both in regard to gap-filling in scrambled sentences (Chapter 2) and in regard to looking for Q-particles (and thereby calculating scope) in mono-clausal and bi-clausal wh-questions (Chapters 3 and 4). Over the past several years, evidence for the incremental nature of Japanese sentence processing has been consistently reported (e.g., Kamide & Mitchell, 1999; Kamide, Altmann, & Haywood, 2003; see Miyamoto, 2003 for an overview), and the present study seems to provide another piece of evidence in support of this point.

As for scrambling, the pattern of results obtained in Chapter 2 is incompatible with a model of sentence processing based on head-driven parsing (Pritchett, 1992). It is possible to say that the slow anterior negative potentials between filler and gap index the effect of the parser’s carrying an extra filler NP before thematic role assignment. However, the phasic LAN and P600 effects around the gap are difficult
to account for under this model, as it predicts no effects of gap-filling before the sentence-final verb.

As for wh-scope calculation, the parser seems to be expecting and actively seeking out a Q-particle following the appearance of a Japanese wh-word well before the Q-particle position, rather than waiting until the Q-particle position. Recall from Chapter 4 that Aoshima, Phillips, and Weinberg (in press) argued for the incremental processing of Japanese wh-elements by showing a Japanese version of a filled-gap effect (Stowe, 1986) before any verb positions. The data by Aoshima et al. and the ERP and reading time evidence shown in Chapters 3-5 of this dissertation all seem to point to the conclusion that the parser actively and incrementally tries to associate a Japanese wh-element with its Q-particle, just like the parser actively tries to associate an English wh-element with its gap position (Active Filler Strategy: Frazier & Clifton, 1989).

6.2.1.3 Overall

Despite marked syntactic differences, language processing mechanisms seem strikingly similar. Japanese and English wh-questions look quite different on the surface, yet both actually utilize mechanisms that license the scope of wh-elements (either by a Q-particle or by displacement) and elicit similar ERP effects ((R)AN vs. (L)AN). When displaced, Japanese wh-elements also elicit processing patterns similar to those seen in wh-gap dependencies in English. The present study, along with recent studies of Japanese sentence processing, point to the incremental processing of Japanese, just as in English, regardless of its verb-final word order.
In a sense, it is sensible that Japanese, despite its syntactic differences, acts in very similar ways to head-initial, wh-movement languages in terms of its parsing operations. The fact that the parsing operations seem to be very similar may help to explain why Japanese speakers do not seem to find the language hard to process after all. It is also interesting to note that languages seem to have various mechanisms to facilitate processing in different ways. Japanese is a verb-final, SOV language, yet it has case-markers that are attached to argument NPs and provide information about the thematic roles and grammatical functions of their NP hosts. This is apparently the case for other SOV languages as well; Greenberg (1966: Universal 41) argues that if a language has predominately SOV order, the language almost always has a case system. Likewise, although the location of Japanese wh-elements does not indicate their scope position, there are Q-particles that perform that function (although this does not always seem to be the case for wh-in-situ languages; see Cheng, 1991).\(^1\)

Yamashita (1994) argues for the universality of a human language processing mechanism, saying that as every human being has approximately the same memory limitations, there must be many aspects of processing that all languages share. Along the same lines, as we have approximately the same neural systems, there should be many aspects of neural processing that different languages share.

---

\(^1\) Cheng (1991) shows that languages that allow for wh-in-situ in single wh-questions invariably use overt morphological Q-markings (such as Q-particles and agreement) in matrix yes/no-questions. However, a majority of wh-in-situ languages in Cheng’s examples (p. 21) do not have overt marking for wh-questions. Cheng argues that in these languages, non-overt marking is present for wh-questions, and that the languages with overt yes-no particles also have wh-particles, be they overt or not. Cheng further proposes that languages with ambiguous wh-words, i.e., wh-words with interrogative, existential and universal interpretations, (e.g., Japanese, Korean) require the presence of overt Q-particles, while languages with non-ambiguous wh-words (e.g., Hindi, Bahasa Indonesia) allow covert Q-particles. Given the above, it would be interesting to study wh-in-situ languages without overt Q-particles in terms of how their wh-scopes are licensed and how their wh-questions are processed.
6.2.2 Question 2: How is the scope of wh-in-situ calculated in real-time processing?

As discussed above, the present study along with Aoshima et al. (in press) suggests that the scope of Japanese wh-in-situ is calculated by the real-time, incremental formation of a long-distance dependency with its corresponding Q-particle.

6.2.3 Question 3: To what extent do neural processing data reflect linguistic theory?

The present study seems to have revealed some possible relationship between linguistic theory and brain responses for both scrambling and wh-processing. With respect to scrambling, the pattern of results obtained in Chapter 2 seems to require some notion of sentence position and syntactic dependency. The slow potentials of AN probably index the working memory load of relating scrambled wh- (and demonstrative) fillers to their canonical or base positions. The P600 probably indexes the effects of syntactic integration at both base and displaced (for the latter, only with scrambled demonstratives) positions, and phasic LAN probably indexes the effects of retrieving fillers from working memory for purposes of syntactic integration. Thus the pattern of results certainly indicates some type of sentence-position-defined dependency that is causing extra processing costs.

In terms of the wh-Q dependency, the processing of both mono-clausal and bi-clausal Japanese wh-questions suggests a reliable neural processing correlate (as indexed by the (R)AN) of the dependency between wh-elements and Q-particles (Chapters 3 and 4). Theoretical claims about wh-in-situ argue for a mechanism inside
COMP that licenses the scope of a corresponding wh-element, either by a Q-particle (Nishigauchi, 1990; Cheng, 1991), or its related Q-operator (Baker, 1970; Pesetsky, 1987; Watanabe, 1992; Aoun & Li, 1993; Cole & Herman, 1994), or the moved wh-element itself (Huang, 1982; Lasnik & Saito, 1984; Nishigauchi, 1990). This is consistent with the present data showing that the brain actively seeks a Q-particle in COMP. From all of the above, we can say that brain responses reflect the complex relationship between a wh-in-situ and its licensing COMP as proposed in various linguistic theories.

Obviously, however, mapping linguistic theories and brain responses is not always a straightforward matter. The first difficulty is the question of which linguistic theory to adopt, with numerous theoretical frameworks and with theories changing within their own frameworks. In the 60’s, a massive volume of research was conducted to test a hypothesis for the psychological reality of grammar called the “derivational theory of complexity (DTC)”, which hypothesized that the complexity of a sentence is a function of the complexity of its derivation. The general conclusion of DTC research was that the structures assigned by transformational grammars are psychologically real, whereas the transformational rules used to generate those structures are not (see Fodor, Bever, & Garrett, 1974 for a review; and see Phillips, 1996 for an objection to this conclusion). However, these experiments were run based on a version of the generative grammar in vogue at the time, and some of the assumptions tested (e.g., reversal of transformations for semantic interpretation, some of the transformational rules) no longer hold in more recent versions of the theory.
The second difficulty is in determining the best way to connect the grammar and the parser. In order to reconcile seeming discrepancies between the grammar and the parser, proposals have been made for modifying either the processing system (Berwick & Weinberg, 1984), the grammar (Bresnan & Kaplan, 1982; Phillips, 1996), or the interpretation of the grammatical representation that interfaces with the processing system (Pritchett & Whitman, 1995).

Related to the above, the third difficulty is how we should map linguistic theories onto word-by-word processing data, or vice versa. Traditionally, linguistic theories have been concerned with a static representation of an entire sentence (although derivational transformations are often assumed) and have not taken dynamic word-by-word processing much into consideration. For instance, X-bar theory (see Section 1.2.1 in Chapter 1) strictly determines hierarchical relationships among syntactic nodes, but does not specify linear relationships. Nevertheless, the direction of Spec projection in a head-final language like Japanese makes an enormous difference in linear word order for word-by-word processing. One way to solve this difficulty is to construct a grammar that derives sentences in a strictly left-to-right fashion, just like the parser would process incoming words (e.g., the PIG (Parser Is Grammar) model by Phillips, 1996).

Going back to the present study, it is certain that more evidence is needed to determine the precise nature of the relationship between the grammar and brain responses. But it seems possible to map the ERP signals obtained in the present study to the related syntactic theories.
6.2.4 Question 4: What is the nature of the relationship (if any) among the data obtained from different methodologies?

The results from the present study revealed that rating, reading time, and ERP measures show a consistent pattern of results with respect to continuing wh-Q dependencies, manifested as the heaviest processing cost to matrix clause wh-questions and the lightest processing cost to embedded clause wh-questions. In addition, the present study suggested that the data from these methodologies generally correlate with each other, with ratings showing the most coarse measures, followed by reading times and then ERPs, but on some occasions different methodologies might reveal different aspects of processing. We have seen one occasion in which ERP measures seem to have revealed sentence-level processing while reading time measures seem to have revealed lexical semantic-level processing.

6.3 Final Conclusions

This dissertation examined the processing of Japanese mono-clausal and bi-clausal wh-questions, using ERPs, reading time measures, and acceptability ratings. The results have demonstrated similarities in how syntactically distinct languages, such as Japanese and English, are processed in the brain, both in terms of ERP components (e.g., LAN/RAN and P600) and incremental processing strategies. The present study along with Aoshima et al. (in press) suggests that the scope of Japanese wh-in-situ is calculated by the incremental formation of a long-distance dependency with its corresponding Q-particle. The present study seems to have revealed some possible relationship between linguistic theory and brain responses for both scrambling and wh-processing. With respect to scrambling, the pattern of results
seems to require some notion of sentence position and syntactic dependency. With respect to wh-scope licensing, brain responses seem to reflect a dependency between a wh-in-situ and its licensing COMP as proposed in various linguistic theories.

Taken all together, the results found in this study seem to suggest that many aspects of language processing, such as syntactic dependencies and neural and behavioral processing mechanisms, are inter-related, and thus interdisciplinary research is an effective approach for teasing them apart. Recall from Chapter 1 that the goal of this research was to bring together theoretical syntax, psycholinguistic processing theories, and cognitive neuroscience. It is hoped that the experimental work in this dissertation has shown some positive movement in this direction, and has demonstrated a promising line of mind/brain research.
Appendix 1: Sample Stimuli for ERP Experiments Discussed in Chapters 2 and 3 and Stimuli for Acceptability Ratings Discussed in Chapters 2, 3, and 5

Experimental Conditions:

a. Wh-in-situ questions

あの/幼い/男の子に/よると/その/奇想天外な/手品師が/いきなり/何を/取出したんですか。

‘According to the young boy, what did the fantastic magician suddenly take out?’

b. Demonstrative(yes/no)-in-situ questions

あの/幼い/男の子に/よると/その/奇想天外な/手品師が/いきなり/それを/取出したんですか。

‘According to the young boy, did the fantastic magician suddenly take that out?’

c. Scrambled wh-questions

あの/幼い/男の子に/よると/何を/その/奇想天外な/手品師が/いきなり/取出したんですか。

‘According to the young boy, what did the fantastic magician suddenly take out?’

d. Scrambled demonstrative (yes/no) questions

あの/幼い/男の子に/よると/それを/その/奇想天外な/手品師が/いきなり/取出したんですか。

‘According to the young boy, did the fantastic magician suddenly take that out?’

Below are the a-versions only.

1a. あの/幼い/男の子に/よると/その/奇想天外な/手品師が/いきなり/何を/取出したんですか。

‘According to the young boy, what did the fantastic magician suddenly take out?’

2a. あの/デパートの/店員に/よると/その/ぜいたくな/社長婦人が/最初に/何を/選んだんですか。

‘According to the department clerk, what did the high-end president’s wife choose first?’

3a. あの/上品な/奥方に/よると/その/後任の/女中さんが/てきぱき/何を/乾かしたんですか。

‘According to the elegant madam, what did the new maid quickly dry?’
4a. あの/とんかつ屋の/経営者に/と/or/その/だらしない/アルバイト学生が/とうとう/何を/返したんですか。

‘According to the owner of the pork cutlet restaurant, what did the sloppy part-time student worker finally return?’

5a. あの/けんもほろろの/役人に/と/or/その/社会派の/ジャーナリストが/堂々と/何を/書いたんですか。

‘According to the arrogant bureaucrat, what did the socialist journalist boldly write?’

6a. あの/担任の/教員に/と/or/その/悪童の/小学生が/こそこそ/何を/隠したんですか。

‘According to the homeroom teacher, what did the wicked elementary school student quietly hide?’

7a. あの/ブティックの/ハウスマヌカンに/と/or/その/大スターの/タレントが/こっそり/何を/買ったんですか。

‘According to the boutique clerk, what did the big star secretly buy?’

8a. あの/ごっついい長男によると/その/可愛らしい/妹が/一生懸命/何を/飾ったんですか。

‘According to the dowdy son, what did his pretty sister decorate fiercely?’

9a. あの/販売部の/営業主任に/と/or/その/不注意な/派遣社員が/またまた/何を/見落したんですか。

‘According to the business manager of the sales department, what did the careless temporary staff miss again?’

10a. あの/大柄な/現場監督に/と/or/その/そそっかしい/作業員が/うっかり/何を/切ったんですか。

‘According to the bulky foreman, what did the unmindful worker cut by mistake?’

11a. あの/近くの/いたずらっ子達によると/その/ブチの/小犬が/偶然/何を/壊したんですか。

‘According to the children of the neighborhood, what did the spotted puppy accidentally break?’

12a. あの/農家の/おかみさんによると/その/乱暴な/少年達が/どさどさ/何を/踏んだんですか。

‘According to the lady of the farm, what did the naughty boys loudly step on?’

13a. あの/秘書室の/みんなによると/その/新参の/秘書が/どうにか/何を/まとめたんですか。

‘According to everyone in the secretarial room, what did the new secretary finally put together?’
14a. あの町内会の副班長によると、そのおとなしい鍵っ子が急に何を見たんですか。

‘According to the vice president of the neighborhood association, what did the quiet child suddenly find?’

15a. あのローカルな新聞によると、その命知らずの冒険家がついに何を発見したんですか。

‘According to the local newspaper, what did the reckless adventurer finally discover?’

16a. あの同室の入院患者によると、その長患いの病人が、あらかじめ何を飲んだんですか。

‘According to a patient in the same room, what did the long-time patient drink in advance?’

17a. あの当直の医師によると、その軽率な糖尿病患者が、知らずに何を食べたんですか。

‘According to the doctor on duty, what did the carefree diabetic patient eat without paying attention?’

18a. あの新興宗教の教団によると、そのいんちきくさい霊媒が、どんどん何を使ったらんですか。

‘According to the new cult organization, what did the phony psychic medium continuously use?’

19a. あのお固い美術委員会によると、その意欲的な芸術家が、一月で何を作ったんですか。

‘According to the conservative art committee, what did the inspired artist create in one month?’

20a. あのライバル支店の課長代理によると、そのすご腕のセールスマンがすんなり何を売ったんですか。

‘According to the deputy manager of the rival branch, what did the skilled salesman sell effortlessly?’
Appendix 2: Sample Stimuli for ERP Experiments Discussed in Chapter 4 and Stimuli for Reading Time Experiment Discussed in Chapter 5 and Acceptability Ratings Discussed in Chapters 4 and 5

Experimental Conditions:

a. Embedded clause wh-questions

専務が/何台の/パソコンを/買ったか/経理課の/係長が/尋ねましたか。

‘Did the manager of the accounting section ask how many PCs the director bought?’

b. Matrix clause wh-questions

専務が/何台の/パソコンを/買ったと/経理課の/係長が/言いましたか。

‘How many PCs did the manager of the accounting section say the director bought?’

c. Embedded clause yes/no-questions

専務が/何台の/パソコンを/買ったか/経理課の/係長が/尋ねましたか。

‘Did the manager of the accounting section ask whether the director bought two PCs?’

d. Matrix clause yes/no-questions

専務が/二台の/パソコンを/買ったと/経理課の/係長が/言いましたか。

‘Did the manager of the accounting section say that the director bought two PCs?’

Below are a- and d-versions only.

1a. 専務が/何台の/パソコンを/買ったか/経理課の/係長が/尋ねましたか。

‘Did the manager of the accounting section ask how many PCs the director bought?’

1d. 専務が/二台の/パソコンを/買ったと/経理課の/係長が/言いましたか。

‘Did the manager of the accounting section say that the director bought two PCs?’

2a. 犯人が/どんなパックを/盗んだか/担当の/刑事が/質問しましたか。

‘Did the detective in charge ask what kind of purse the criminal stole?’

2d. 犯人が/ブランド物の/パックを/盗んだと/担当の/刑事が/報告しましたか。
‘Did the detective in charge report that the criminal stole a brand-name purse?’

3a. 园児達が/どんな/野菜を/食べ残したか/保育園の/保母さんが/聞きましたか。

‘Did the kindergarten teacher ask what kind of vegetables the children left?’

3d. 园児達が/かたい/野菜を/食べ残したと/保育園の/保母さんが/言いましたか。

‘Did the kindergarten teacher say that the children left hard vegetables?’

4a. お客さんが/何杯の/コーヒーを/頼んだか/喫茶店の/店長が/尋ねましたか。

‘Did the coffeeshop manager ask how many cups of coffee the customer ordered?’

4d. お客さんが/二杯の/コーヒーを/頼んだと/喫茶店の/店長が/話しましたか。

‘Did the coffeeshop manager say that the customer ordered two cups of coffee?’

5a. 会長が/どんな/歌を/歌ったか/現場の/職員が/問い掛けましたか。

‘Did the staff on site ask what kind of song the chairperson sang?’

5d. 会長が/流行の/歌を/歌ったと/現場の/職員が/ささやきましたか。

‘Did the staff on site whisper that the chairperson sang a popular song?’

6a. 教授が/どんな/本を/出版したか/研究室の/助手が/思い出しましたか。

‘Did the research assistant remember what kind of book the professor published?’

6d. 教授が/雑学の/本を/出版したと/研究室の/助手が/つぶやきましたか。

‘Did the research assistant murmur that the professor published a book of trivia?’

7a. サッカー選手が/何年の/契約を/交わしたか/番組の/レポーターが/質問しましたか。

‘Did the reporter from the program ask how many years the soccer player signed a contract for?’

7d. サッカー選手が/三年の/契約を/交わしたと/番組の/レポーターが/報告しましたか。

‘Did the reporter from the program report that the soccer player signed a three-year contract?’
8a. 部長が「どんな/機械を/壊したか/現場の/工事長が/確認しましたか。」
‘Did the factory manager confirm what kind of machine the director broke?’

8d. 部長が「大切な/機械を/壊したと/現場の/工事長が/断言しましたか。」
‘Did the factory manager claim that the director broke an important machine?’

9a. 評論家が「何人の/作家を/誉めたか/文芸雑誌の/編集長が/尋ねましたか。」
‘Did the chief editor of the literary magazine ask how many authors the critic praised?’

9d. 評論家が「二人の/作家を/誉めたと/文芸雑誌の/編集長が/語りましたか。」
‘Did the chief editor of the literary magazine say that the critic praised two authors?’

10a. お手伝いさんが「どんな/お皿を/割ったか/お屋敷の/奥さんが/問詰めましたか。」
‘Did the lady of the house demand to know what kind of dish the maid broke?’

10d. お手伝いさんが「大切な/お皿を/割ったと/お屋敷の/奥さんが/決付けましたか。」
‘Did the lady of the house insist that the maid broke an important dish?’

11a. 老人が「何万円の/現金を/なくしたか/派出所の/警察官が/調査しましたか。」
‘Did the police officer at the police station investigate how many tens of thousands of yen the elderly person lost?’

11d. 老人が「五万円の/現金を/なくしたと/派出所の/警察官が/書き留めましたか。」
‘Did the police officer at the police station write down that the elderly person lost ¥50,000?’

12a. 魚屋さんが「どんな/魚を/仕入れたか/町内の/お寿司屋さんが/聞出しましたか。」
‘Did the sushi chef of the town inquire what kind of fish the fish shop purchased?’

12d. 魚屋さんが「新鮮な/魚を/仕入れたと/町内の/お寿司屋さんが/口外しましたか。」
‘Did the sushi chef of the town leak that the fish shop purchased fresh fish?’
13a. 小学生が何冊の本を読んだかアルバイトの塾講師が尋ねましたか。

‘Did the part-time cram school teacher ask how many books the elementary school student read?’

13d. 小学生が二冊の本を読んだとアルバイトの塾講師が伝えましたか。

‘Did the part-time cram school teacher report that the elementary school student read two books?’

14a. 新入社員がどんな書類をなくしたか総務部の事務員が間詰めましたか。

‘Did the clerk of the administrative department demand to know what kind of documents the new employee lost?’

14d. 新入社員が重要な書類をなくしたと総務部の事務員が口外しましたか。

‘Did the clerk of the administrative department leak that the new employee lost important documents?’

15a. 作業員が幾つの箱を運んだか当直の係員が確認しましたか。

‘Did the person on duty confirm how many boxes the workers carried?’

15d. 作業員が四つの箱を運んだと当直の係員が証言しましたか。

‘Did the person on duty testify that the workers carried four boxes?’

16a. 息子が何人の女性を愛したか田舎の母親が気付きましたか。

‘Did the mother in the country notice how many women her son loved?’

16d. 息子が二人の女性を愛したと田舎の母親が打明けましたか。

‘Did the mother in the country confess that her son loved two women?’

17a. 秘書がどんな資料を捨てたか役員室の室長が調査しましたか。

‘Did the chief of the executive assistant room investigate what kind of documents the secretary threw away?’
17d. 秘書が極秘の資料を捨てたと役員室の室長が主張しましたか。

‘Did the chief of the executive assistant room insist that the secretary threw away the confidential documents?’

18a. おばあさんが何本のバラを植えたか近所の孫が聞きましたか。

‘Did the grandchild in the neighborhood ask how many bunches of roses his grandmother planted?’

18d. おばあさんが三本のバラを植えたと近所の孫が言いましたか。

‘Did the grandchild in the neighborhood say that his grandmother planted three bunches of roses?’

19a. 一年生がどんな花瓶を割ったか中学校の教頭先生が見破りましたか。

‘Did the vice-principal of the junior high school figure out what kind of vase the first-year student broke?’

19d. 一年生が高価な花瓶を割ったと中学校の教頭先生が決め付けましたか。

‘Did the vice-principal of the junior high school determine that the first-year student broke the expensive vase?’

20a. 生徒が幾つの選択科目を勉強したか臨時雇いの家庭教師が確認しましたか。

‘Did the temporary tutor confirm how many electives the student studied?’

20d. 生徒が三つの選択科目を勉強したと臨時雇いの家庭教師が報告しましたか。

‘Did the temporary tutor report that the student studied three electives?’

21a. 課長が何軒の家を建てたか人事課の部下が聞きましたか。

‘Did the subordinates in the human resources department ask how many houses the section manager built?’

21d. 課長が二軒の家を建てたと人事課の部下が噂しましたか。

‘Did the subordinates in the human resources department gossip that the section manager built two houses?’
22a. 同級生がどんな服を見せびらかしたか/クラスの女子高生が尋ねましたか。

‘Did the high school girls in the class ask what kind of clothes their classmate showed off?’

22d. 同級生が派手な服を見せびらかしたとクラスの女子高生が噂しましたか。

‘Did the high school girls in the class gossip that their classmate showed off flashy clothes?’

23a. アイドルが何冊の写真集を出したか/ファンクラブの会員が質問しましたか。

‘Did the member of the fan club question how many volumes of photo books the pop singer published?’

23d. アイドルが二冊の写真集を出したとファンクラブの会員が自慢しましたか。

‘Was the member of the fan club proud that the pop singer published two volumes of photo books?’

24a. 女性達がどんな料理を作ったか/同僚の男性達が聞きましたか。

‘Did the male coworkers ask what kind of dish the women cooked?’

24d. 女性達がインド料理を作ったと同僚の男性達が言いましたか。

‘Did the male coworkers say that the women cooked an Indian dish?’
REFERENCES


Kluender, R., & Münte, T. F. (1998). ERPs to grammatical and ungrammatical wh questions in German: Subject/object asymmetries. Poster session presented at the 11th annual CUNY Conference on Human Sentence Processing, Newark, NJ.


