EFFECTS OF SIMULATION IN SENTENCE PRODUCTION

A DISSERTATION SUBMITTED TO THE GRADUATE DIVISION OF
THE UNIVERSITY OF HAWAI‘I IN PARATIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

LINGUISTICS

MAY 2010

By
Manami Sato

Dissertation Committee:

Amy J. Schafer, Chairperson
Benjamin K. Bergen
William O’Grady
Hiromu Sakai
Zhang Shuqian
ACKNOWLEDGMENTS

I cannot count how many times I have dreamt about writing acknowledgments as the final touch of the dissertation completion during this great journey as a Ph.D. student in the Department of Linguistics at the University of Hawai‘i. Sitting in my favorite studying spot (where I can comfortably focus on writing) and watching Diamond Head against the clear blue sky, I’m thinking of my Ph.D. journey that has been enriched by invaluable advice and continuous support from faculty members and colleagues and nourished by the crazy laughter my friends brought. No words can express enough my appreciation to those who guided me or walked beside me unconditionally.

First and foremost, I would like to express my deepest gratitude and respect to my advisor and chair of the dissertation committee, Amy J. Schafer, for her sharp and innovative mind and eternal support. She has trained me to become an independent researcher in the future (although it will be a long way to truly become a researcher like her) and taught me the richness and excitement of psycholinguistic study. Under her enthusiastic guidance and sparkling inspiration, I have learned to critically assess the existing studies, dynamically postulate innovative questions without setting a boundary between different fields of study, empirically investigate the questions, and extract the theoretical findings from the results. I am very grateful for the extensive time and knowledge she has generously shared with me (even when it has meant that she had to sacrifice her lunch time or weekends) and for her continuous encouragement and trust (even during the period of time when I couldn’t trust myself). Her thorough and critical feedback on conference abstracts, conference presentations, grant applications, manuscript submissions, and this dissertation for past years encouraged me to move
forward and gradually shaped who I am now academically and personally. I would also like to thank her for her open mind and great capacity to supervise research on mental simulation in sentence production. A positive, warm, and welcoming atmosphere that comes from her innate characteristics as well as her enthusiasm about research have been influential to many students from various departments and always stimulated thought, inspired me to produce new projects, and gave me a fruitful period of time to complete the dissertation. I will follow her footsteps one by one.

I would like to thank Benjamin K. Bergen for teaching me cognitive linguistic research in the most astonishing way possible. His energetic and entertaining courses always made me forget that I was in class, and my intellectual curiosity had awakened to the intricacies of mental simulation before I realized it. I have learned tremendously from his knowledge (in class, and also in places where we could all have beers opened with his famous “abused” bottle opener) and I felt very lucky that he was the one who introduced me to the field and led me to the entrance of my research journey exploring how language and mind work.

My sincere gratitude also goes to William O’Grady, who has provided significant advice on the theory of event representations and word order. I would like to thank him especially for his prompt and substantial feedback even when his schedule was extremely tight (like the very next day after returning from Hong Kong).

I would like to thank Hiromu Sakai (Hiroshima University), who has generously provided guidance on the conception of sentence production in this work. Furthermore, he welcomed me to Hiroshima University for the data collection for this study and previous research while I was working as a research fellow at the Center for Brain
Science of Language Acquisition and Language Learning (CBL) at Hiroshima

University, and provided me opportunities to present my work that helped me to improve the study. His sharpness and high productivity as a scholar and his hospitality and warmness as a person have clearly affected the people around him, and I was very surprised and delighted to see his influence on his students and colleagues.

I am also thankful for Zhang Shuqiang’s generous support and help on the statistical analyses. I am grateful to have met my two honest and precise research assistants, Mariya Hayakawa and Jun Endo, who selflessly devoted their time for coding and checking the tremendous amount of data. I am also grateful to Katsuo Tamaoka (Nagoya University) for inviting me to present this study. My appreciation also goes to Laurie Durand, who has provided me immediate and fruitful editorial support to improve this work by showing me various ways to present a single concept. I would also like to thank our extraordinary secretaries, Jennifer Kanda and Nora Lum, who have welcomed me every morning and who have precisely but warmly assisted me throughout the Ph.D. life.

Completing this research would not have been possible without the financial support of several organizations. I would like to acknowledge the National Science Foundation: Dissertation Research Improvement Grants, the Department of Linguistics at the University of Hawai‘i for five Endowment Fund grants, the Graduate Student Organization at the University of Hawai‘i for a research grant, two Dai Ho Chun Scholarships, the Office of Community and Alumni Relations at the University of Hawai‘i for an Arts and Sciences Advisory Council Award, the Graduate Student Travel Grant from the Center for Japanese Studies at the University of Hawai‘i, and travel grants
from three conferences I have attended: JSA-ASEAN Travel Grant, Mental Architecture for Processing and Learning of Language (MAPLL), the CUNY Graduate School & University Center.

Throughout the Ph.D. program, I have gained precious friends and lost some. I would like to send my most sincere and deepest love to Poco, who has shown and taught me priceless loyalty and unconditional emotional attachment for seven years. Countless adventurous memories with you in Santa Rosa, Healdsburg, Berkeley, Sapporo, Tokyo, and Hawai‘i will never fade, and you will keep living in me as I promised. I would also like to express my gratitude to my friend, Dan X. Hall, who never gave up and kept fighting with the unbeatable one with a positive spirit, and who wanted to see me complete the Ph.D. so badly, but couldn’t wait.

In Hawai‘i, I have met a non-genetic “sister,” Mie Hiramoto, who has been a model of how to be a successful scholar as well as how to be an extraordinarily attractive person who continuously produces positive energy and provides unlimited care and support. Her consistent convictions toward research and her joyful characteristics have heavily influenced my way of thinking. Likewise, I am thankful to Benjamin George for his heartfelt support.

I am indebted to many friends for their support and kindness. I especially thank the gang members, Tracy (Apay, Ai-yu) Tang and Soyoung Kim, for their never changing friendships. In difficult times and delightful times, we have cried and laughed together, as if we were still kids. Although there seemed to be no solution for the problems we were faced with, we gathered and never abandoned each other: instead, we chose to laugh together. Their selfless support and thoughtfulness made my life in
Hawai‘i colorful and unforgettable. I would also like to thank Yukie Hara for studying and chatting together many nights (and sneaking out for refreshing our minds), Yumiko Enyo for her generous understanding and delicate kindness, Nian Liu and Hao for their cheerful attitudes (and for experiencing many brave adventures in Kyoto and Osaka), Heeyeon Dennison for experiencing the busiest time and unexpected hardships in the Ph.D. program together (and finally exploding and releasing our stress in NYC), Hyekyung Hwang for her considerate heart and care. My appreciation similarly goes to Sachie Jansen, Hye-Young Kwak, Onsoon Lee, Kaori Ueki, Toshiaki Furukawa, Jun Nomura, Yohe Sakai, Jawee Perla, Carl Polley, Kathryn Wheeler, Bodo Winter, Jin-Sook Kim, Chae Eun Kim, Jung Hee Kim, Jin Sun Choe, Hiroko Sato, Mari Miyao, Elaine Lau, Antonio Cheung, Karen Huang, Suann Robinson, Hunter Hatfield, and Kum Jeong Joo.

I am truly grateful to my parents, Hisao and Seiko Sato, who have been unconditionally giving me their overwhelming support and love throughout my life. They have always respected and supported my interests and goals, and have entirely cherished my progress more than I do. Moreover, I felt very lucky to be raised by an inseparable, lovely couple who actively enjoy daily morning walks, gardening, going to concerts, and traveling the world together. Dad, I am very thankful to you for protecting us and providing us the best in all aspects with your generosity and dignity of love. My best friend with a sparkling spirit, Mom, you are full of joyful surprises – I am looking forward to learning hula, belly dance, yoga, and flower arranging, just to name a few, from you. Shine on us as usual. And my brother, Keisuke Sato, because you were there, I have been able to pursue my goal.
Last but not least, I am deeply indebted to Ken (among many other nicknames) for happily accepting exploring this journey with me from the very beginning. Every single experience we have gone through together at different stages of our 14 years-long adventurous exploration is precious and just thinking of him always gives me a smile and peace of mind. His humorous personality makes me laugh (so hard that I immediately forget why I was down), his trustworthy heart and everlasting love make me stronger, and his vitality and professionalism amaze me and inspire me. Without him, I could not have taken this journey.

Of course, I am entirely responsible for any errors and misunderstandings that might be found in this dissertation.
ABSTRACT

This study investigates the role of mental simulation in message formulation and grammatical encoding in two typologically distinct languages, English and Japanese. It examines relationships among physical motion, mental simulation, and sentence production, following the claims of Perceptual Symbol Systems (Barsalou, 1999) that people understand language by mentally simulating multimodal experiential knowledge, and that such simulation involves activation of our sensorimotor systems (Barsalou, 2007; Gallese, 2007). Specifically, it assesses whether speakers’ embodied status, manipulated by motor activities, can influence message choice and word order.

The relationship between motion and language production mechanisms was examined with four factors: language (English vs. Japanese), direction of prime action (toward-, away-, or neutral-motion), timing between motion and message formulation (motion before vs. after the onset of message formulation), and message status (“unconstrained message” vs. “constrained message”). An unconstrained message was one for which the relational meanings of two objects were under-determined: participants saw pictures of two objects and described an action involving those two objects. In contrast, in a constrained message, participants described a fixed event depicted by a sequence of pictures.

The results reveal that regardless of language and message status, motion has an impact on the message planning process. In unconstrained utterances, motion drives people to produce sentences with a corresponding directional orientation: participants produced more “toward sentences” that implied movement toward the speaker’s body
after toward motions, while they produced more “away sentences” that implied
movement away from their bodies after away motions, compared to those after neutral
motions. Furthermore, the results in Japanese show that participants favor the word order
that presents referents in an order consistent with the corresponding situation described in
the sentence. Crucially, motion boosts the relationship between event language and word
order.

These results indicate that speakers are fundamentally responsive to embodied
information that is activated through physical motion, regardless of presence or absence
of relational meanings in the intended message, and regardless of the timing. Moreover,
on-line, causal effects of simulation on event language and word order determination
suggest that simulation is a critical part of our language that is cognitively grounded in
embodied experiences.
TABLE OF CONTENTS

Acknowledgments.............................................................................................................. iii
Abstract................................................................................................................................ ix
List of Tables......................................................................................................................xiv
List of Figures.................................................................................................................... xv

Chapter 1: Introduction.....................................................................................................1

Chapter 2: Mental simulation in comprehension.............................................................8
  2.1 Perceptual symbol systems .................................................................................... 9
  2.2 Perceptual simulation of non-motion features in comprehension............ 12
  2.3 Perceptual simulation of motion features in comprehension.................... 17
    2.3.1 Timing aspects of conceptual processing and sensorimotor processing ......................................................... 18
    2.3.2 Language-mediated perceptual simulation of motion features.... 19
      - Match “facilitation” in sequential processing of simulation and visual stimuli ............................................ 19
      - Match “inhibition” in simultaneous processing of simulation and visual stimuli ........................................ 21
      - Match “inhibition” in a visual field overlap between simulation and an unrelated object .............. 22
  2.4 Language-mediated motor simulation: Action Compatibility Effect (ACE) ...................................................................................... 24
  2.5 Physical information influences simulation during sentence comprehension ........................................................................ 28

Chapter 3: Mental simulation in production.................................................................31
  3.1 Impediments inherent in research on simulation in production................. 31
  3.2 Sentence production model .............................................................................. 33
  3.3 Possible effect of external factors on thought and speech ...................... 36
  3.4 Syntactic priming as a tool to illuminate the relationships between conceptual and grammatical encodings in the production process ........ 37
  3.5 Word order in simulation and non-linguistic factors that influence linguistic construction ........................................................................ 41
  3.6 Gestures as an indicator of simulation during speech production .......... 50
    3.6.1 Functional role of gestures in speech ...................................................... 51
    3.6.2 Gestures as Simulated Action (GSA) ................................................. 52
    3.6.3 Limitations underlying gesture-in-speech studies .............................. 55

Chapter 4: Clock task......................................................................................................59
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1</td>
<td>Apparatus</td>
<td>59</td>
</tr>
<tr>
<td>4.2</td>
<td>Purpose of using the clock-patterned mouse pad</td>
<td>60</td>
</tr>
<tr>
<td>4.3</td>
<td>Experiment setup</td>
<td>60</td>
</tr>
</tbody>
</table>

**Chapter 5: Simulation in unconstrained message formulation** |

5.1 Motion *before* formulating an unconstrained message | 64 |
5.1.1 Participants (English, Japanese) | 64 |
5.1.2 Materials (English, Japanese) | 64 |
5.1.3 Procedure | 68 |
5.1.4 Conditions | 70 |
5.1.5 Data analysis | 71 |
5.1.6 Predictions | 75 |
5.1.7 Results | 76 |
5.1.7.1 English speakers (Experiment 1) | 77 |
5.1.7.2 Japanese speakers (Experiment 2) | 80 |
5.1.8 Discussion | 85 |

5.2 Motion *while* formulating an unconstrained message | 88 |
5.2.1 Participants (English, Japanese) | 89 |
5.2.2 Materials (English, Japanese) | 89 |
5.2.3 Procedure | 89 |
5.2.4 Conditions | 90 |
5.2.5 Data analysis | 91 |
5.2.6 Predictions | 92 |
5.2.7 Results | 93 |
5.2.7.1 English speakers (Experiment 3) | 94 |
5.2.7.2 Japanese speakers (Experiment 4) | 95 |
5.2.8 Discussion | 98 |

5.3 General discussion | 99 |

**Chapter 6: Simulation in fixed message production** |

6.1 Motion *before* formulating a fixed message | 106 |
6.1.1 Participants (English, Japanese) | 106 |
6.1.2 Materials (English, Japanese) | 107 |
6.1.3 Procedure | 110 |
6.1.4 Conditions | 112 |
6.1.5 Data analysis | 113 |
6.1.6 Predictions | 116 |
6.1.7 Results | 118 |
6.1.7.1 English speakers (Experiment 5) | 118 |
6.1.7.2 Japanese speakers (Experiment 6) | 120 |
6.1.8 Discussion | 124 |

6.2 Motion *while* formulating a constrained message | 126 |
6.2.1 Participants (English, Japanese) | 127 |
Chapter 7: General discussion: Action-In-Message (AIM) effects

7.1 Motion, embodied cognition, and message formulation .............148
7.2 Implications .........................................................................150
7.3 Limitations .........................................................................154
7.4 Conclusion .........................................................................155

Appendix A: Experiments 1-4, Critical pictures .................................158
Appendix B: Experiments 2 & 4 (Japanese), Filler pictures .................161
Appendix C: Experiments 1 & 3 (English), Filler pictures ....................162
Appendix D: Experiments 1 & 3 (English), Norming for pictures ..........165
Appendix E: Experiments 2 & 4 (Japanese), Norming for pictures ..........175
Appendix F: Critical pictures in Experiments 5-8 .................................182
Appendix G: Filler pictures in Experiments 5-8 .................................189

References ............................................................................192
## LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5-1</td>
<td>Number of productions with Away, Toward, Neutral, or Other responses in the motion-first, picture-second design</td>
<td>77</td>
</tr>
<tr>
<td>5-2</td>
<td>Sample utterances for each type of language category in Experiment 1</td>
<td>79</td>
</tr>
<tr>
<td>5-3</td>
<td>Sample utterances for each type of language category in Experiment 2</td>
<td>81</td>
</tr>
<tr>
<td>5-4</td>
<td>Number of productions with Away, Toward, Neutral, or Other responses produced after picture-first, motion-second design</td>
<td>93</td>
</tr>
<tr>
<td>6-1</td>
<td>A sample pair of pictures that depict toward- or away-events</td>
<td>108</td>
</tr>
<tr>
<td>6-2</td>
<td>RTs of speech onset with Away or Toward responses in motion-first, constrained-message-second design</td>
<td>118</td>
</tr>
<tr>
<td>6-3</td>
<td>Number of responses with canonical word order, scrambled word order, or other word order in Japanese in motion-first, constrained-message-second design</td>
<td>123</td>
</tr>
<tr>
<td>6-4</td>
<td>RTs for mouse click with Away or Toward event descriptions</td>
<td>134</td>
</tr>
<tr>
<td>6-5</td>
<td>RTs for speech onset with Away or Toward event descriptions</td>
<td>134</td>
</tr>
<tr>
<td>6-6</td>
<td>Number of responses with canonical word order, scrambled word order, or other word order in Japanese in constrained-message-first, motion-second design</td>
<td>139</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>------------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>3-1</td>
<td>Model of sentence production (adapted from Bock &amp; Levelt, 1994; Ferreira &amp; Slevc, 2007; Levelt, 1999)</td>
<td>35</td>
</tr>
<tr>
<td>4-1</td>
<td>A large clock-patterned mouse pad</td>
<td>59</td>
</tr>
<tr>
<td>4-2</td>
<td>Configuration of apparatus</td>
<td>61</td>
</tr>
<tr>
<td>5-1</td>
<td>Experiment 1 (English). Away motion followed by a sequence of an object picture and a location picture</td>
<td>69</td>
</tr>
<tr>
<td>5-2</td>
<td>Experiment 2 (Japanese). Toward motion followed by a sequence of a location and an object picture</td>
<td>69</td>
</tr>
<tr>
<td>5-3</td>
<td>Percentage of English responses after performing motion followed by picture perception</td>
<td>79</td>
</tr>
<tr>
<td>5-4</td>
<td>Percentage of Japanese responses after performing motion followed by picture perception</td>
<td>81</td>
</tr>
<tr>
<td>5-5</td>
<td>Counts and percentages of Toward and Away responses with canonical or scrambled word order after performing motion in specific directions (Experiment 2)</td>
<td>85</td>
</tr>
<tr>
<td>5-6</td>
<td>Experiment 3 (English). A sequence of an object and a location picture followed by left-motion</td>
<td>90</td>
</tr>
<tr>
<td>5-7</td>
<td>Experiment 4 (Japanese). A sequence of a location and an object picture followed by right-motion</td>
<td>90</td>
</tr>
<tr>
<td>5-8</td>
<td>Percentage of English responses after perceiving pictures followed by motion</td>
<td>95</td>
</tr>
<tr>
<td>5-9</td>
<td>Percentage of Japanese responses after perceiving pictures followed by motion</td>
<td>96</td>
</tr>
<tr>
<td>5-10</td>
<td>Counts and percentages of Toward and Away responses with canonical or scrambled word order after performing motion in specific direction (Experiment 4)</td>
<td>97</td>
</tr>
<tr>
<td>6-1</td>
<td>Experiment 5 (English). Away-motion followed by away-message</td>
<td>111</td>
</tr>
</tbody>
</table>
6-2 Experiment 6 (Japanese). Toward-motion followed by toward-message……..112

6-3 Speech initiation times (milliseconds) of toward- and away-language after match, neutral, or mismatch motion in Experiment 5……………………………………..119

6-4 Speech initiation times (milliseconds) of toward- and away-language after match, neutral, or mismatch motion in Experiment 6……………………………………..120

6-5 Counts and percentages of toward and away responses with canonical or scrambled word order after performing motion in specific directions (Experiment 6)…………………………………………………………………………….124

6-6 Experiment 7 (English). Away-message followed by left-motion………………128

6-7 Experiment 8 (Japanese). Toward-message followed by right-motion……….128

6-8 Speech initiation times (milliseconds) of toward- and away-language after match, neutral, or mismatch motion in Experiment 7…………………………………...136

6-9 Speech initiation times (milliseconds) of toward- and away-language after match, neutral, or mismatch motion in Experiment 8……………………………………..138

6-10 Counts and percentages of Toward and Away responses with canonical or scrambled word order after performing motion in specific directions (Experiment 8)……………………………………………………………………..140
CHAPTER 1
INTRODUCTION

How do speakers generate messages? Where and how do these messages originate, and what influences message formulation in speakers’ minds? What cognitive processes do speakers experience when transferring prelinguistic messages into linguistic components? These are inspiring yet largely unexplored questions within the fields of psycholinguistics and cognitive linguistics, and this study aims to contribute to research on these essential topics.

Meaning transfer in language comprehension is often studied within the framework of simulation semantics in cognitive linguistics. This framework focuses on the question of what enables listeners to successfully understand and unpack speakers’ intended messages. For example, Barsalou (1999) and others have argued that the successful transfer of meaning between speakers and listeners is due to perceptual and motor simulation, i.e., listeners mentally create or recreate real-world events when listening to speech about these events. This process then activates recurrent patterns of neural pathways of our sensorimotor systems, enabling listeners to actually experience speakers’ conveyed meanings. The importance of this mental simulation has been recognized for a decade (Barsalou, 1999), and behavioral studies have produced a growing body of evidence indicating that people employ mental simulation during language comprehension in conjunction with understanding a described meaning. Specifically, researchers have found that listeners simulate implied object properties (Stanfield & Zwaan, 2001; Zwaan, Stanfield, & Yaxley, 2002; Zwaan & Yaxley, 2004), motor
properties (Glenberg & Kaschak, 2002; Kaschak et al., 2005; Setti, Borghi, & Tessari, 2009), and intrinsic spatial information denoted by nouns (e.g., sky) and verbs (e.g., climb) (Bergen, Lindsay, Matlock, & Narayanan, 2007; Estes, Verges, & Barsalou, 2008; Meier & Robinson, 2004). Moreover, language can induce simulations involving motion in a specific direction, such as toward-motion (e.g., Open the drawer) or away-motion (e.g., Close the drawer) (Glenberg & Kaschak, 2002; Kaschak et al., 2005). This simulation occurs because language meaning is grounded in our experiences. During our interactions with the world, we store and generalize our perceptual and motor experiences. Reactivation of this stored information during language comprehension results in perceptual or motor simulation of the described scene.

The conclusions drawn from these behavioral studies are supported by neural imaging research that demonstrates that simply imagining specific bodily movements, such as hand-, foot-, and mouth-related actions, systematically recruits the same spatial section of the brain as actually executing these same motions would (Ehrsson, Geyer, & Naito, 2003; Pulvermüller, Haerle, & Hummel, 2001; Tettamanti et al., 2005). This supports the claim that language comprehenders understand language about physical actions by retrieving the experiential-based knowledge they have gained through previous perceptual and motor activities, and use this knowledge to run mental simulations about these physical actions. In other words, simulation is not the product of imagery encapsulated within an individual word in a fixed manner, but rather the dynamic product of imagery induced by integrating sentential meaning with past experience.

However, there are obvious gaps that need to be filled in the research on simulation and language. Previous research investigating the role of simulation deals only with
comprehension or evidence from gestures. Studies focusing on mental simulation in language production are very limited in both number and in the methodology they employ compared to studies on simulation in comprehension.

Research on production mechanisms has been largely the domain of psycholinguistics. Although various production models propose explanations of how non-linguistic conceptual materials are delivered to the subsequent grammatical and phonological encoding processes, the underlying assumption across all models is that production begins when speakers consciously formulate a prelinguistic concept of their intended message (Bock, 1995; Bock & Levelt, 1994; Ferreira & Slevc, 2007; Levelt, 1999). However, this assumption does not address the question of whether speakers always have control over how they frame the message, or whether they are unconsciously influenced by their surrounding environment or their current cognitive status.

The intuitive answer is that speakers are indeed influenced by these factors, based on three considerations. First, because people are constantly engaged in various physical activities while accessing and retrieving knowledge, the language processing mechanisms may also dynamically integrate physical motion into the speaker’s cognitive states and finally into generation of the speaker’s intended meaning. Second, experiential-based knowledge (that is, knowledge that people gain through interaction with particular objects in their surrounding environment) is closely coupled with language because we acquire many of the concepts or meanings encoded in language through our perceptual and motor experiences (Barsalou, 1999; Kaschak & Glenberg, 2000; Tucker & Ellis, 1998; Valenti & Costall, 1997; Yeh & Barsalou, 2006). Language use therefore activates these associated aspects of speakers’ experiential-based knowledge. Conversely, due to this
tight bond between experiential knowledge and language, performing particular actions may automatically activate related concepts within the speakers’ accumulated knowledge base, which plays a facilitative role in language production. Finally, we are part of an existing environment and are never detached from the external world; therefore, thoughts could emerge not solely from our intention, but also from our concurrent, non-linguistic cognitive and bodily status. In fact, previous research in psycholinguistics and cognitive linguistics has implied that non-linguistic factors such as comprehenders’ spatial environments (Boroditsky, 2000; Boroditsky & Ramscar, 2002; Casasanto & Boroditsky, 2008) and their initial visual attention (Gleitman, January, Nappa, & Trueswell, 2007) can unconsciously influence language comprehension and production. Based on these three reasons, it is logical to assess bodily motion as one of the potential non-linguistic factors that may interact with language processing and to investigate whether speakers are sensitive to momentarily salient information from their bodily motion when the speakers are generating both constrained and unconstrained messages.

Then, how can we empirically investigate whether speakers’ current physical/cognitive status unconsciously influences production mechanisms involving their unconstrained as well as constrained messages? I propose that we can empirically investigate this topic by studying mental simulation and external factors such as gesturing (or more specifically, moving a hand toward a certain direction) during language production. The current study builds on both the previous comprehension-oriented simulation studies and Bock and Levelt’s (1994) well-accepted language production model to explore the challenging but important question of whether or not simulation and embodied cognition are influential components of our language production mechanism.
The experiments conducted in this research and described in this dissertation extend the notion of the Action Compatibility Effect (ACE) (ACE, Glenberg & Kaschak, 2002; see section 2.3.4) to sentence production and utilize it as a tool for investigating the relationship between non-linguistic body motion and the meaning construction process. The experiments for this study investigate mental simulation/embodied cognition using evidence from choice of linguistic content, word order, and speech onset time during sentence production in English and Japanese. The inclusion of Japanese, a language with relatively flexible word order, allows critical insight into simulation in sentence production because word order can reflect the temporal order of the simulated event (O’Grady & Lee, 2005; O’Grady, Yamashita, & Lee, 2005). If simulation is part of language production, speakers should show a preference for utterances in which the linguistic arguments are in an isomorphic/chronological order corresponding to their roles in the described event. In addition, comparisons between these two languages can provide insight into whether the role of motor actions in language production remains the same regardless of different linguistic features. If this is proven to be the case, the phenomenon could more plausibly be considered language-independent.

This chapter briefly describes the theoretical background and empirical foundations that inspired this study. Chapters 2 and 3 then discuss the existing theoretical and empirical research in language comprehension and production, respectively. Specifically, Chapter 2 presents the Perceptual symbol systems, which are the backbone of the concept of mental simulation, and outlines cognitive linguistic and neuroimaging research on mental simulation in language comprehension. This chapter also includes a discussion of the core underlying notion of the dissertation, the Action Compatibility Effect. Chapter 3
discusses the reasons why very little research on simulation in language production has been conducted up until this point, and then briefly presents key sentence production models. This chapter also introduces gesture studies, which have provided valuable, but limited, evidence for embodied simulation in language production.

The second part of the dissertation describes the current research. Chapter 4 illustrates the novel “clock task” that was employed throughout all the experiments conducted for this dissertation. This task was created to investigate the active use of simulation in language formulation. Chapter 5 describes a set of four experiments (Experiments 1–4) that investigates the relationship between motor actions and under-determined message formulation processes in the production of English and Japanese utterances. An under-determined message refers to a message wherein some individual concepts (e.g., plug and outlet) are activated, but the details of their relationship are not yet encoded, e.g., the speaker must generate the relational meaning in order to frame the event or the message. Chapter 6 describes another set of four experiments (Experiments 5–8) that aims to examine the relationship between motor actions and messages with fixed internal structures (thus allowing for very little linguistic variation). These two sets of experiments elucidate the role of non-linguistic physical information in language production and show how powerfully this information can manipulate or interact with meaning generation (Experiments 1–4) or meaning extraction/description (Experiments 5–8). Finally, Chapter 7 discusses the implications of the current findings and proposes a language production model that incorporates non-linguistic physical information.

There are many previous studies that empirically show clear effects of language on motion when a motion or bodily state was involved in the critical response in sentence
comprehension (e.g., match effects between sentential meaning and subsequent motor action) (Glenberg & Kaschak, 2002; Zwaan & Taylor, 2006). However, to my knowledge, this is the first study that investigates the converse: that is, the effects of motion on language (e.g., how motor action dynamically changes subsequent linguistic output). This study therefore seeks to make an important and unique contribution to the field of psycholinguistics and cognitive linguistics by showing a new kind of internal connection between language formulation and non-linguistic experience.
CHAPTER 2
MENTAL SIMULATION IN COMPREHENSION

Thoughts can be successfully transferred from speakers or writers to listeners or readers because we often have indirect access to the behaviors of others by using our own multimodal experiential knowledge (Gallese, 2007). A vast body of behavioral research has empirically supported this embodiment view of language understanding, which proposes that simulation and embodied cognition play a central role in representing perceptual (Stanfield & Zwaan, 2001; Zwaan, Stanfield, & Yaxley, 2002; see section 2.2) or motor and spatial language meaning (Glenberg & Kaschak, 2002; see section 2.3). Although various stances exist among all theories of Grounded Cognition regarding whether mental simulation is an obligatory aspect of language processing or whether it simply supplements our language comprehension abilities (Bergen, 2007; Zwaan, 1999; Zwaan & Madden, 2005; Zwaan & Radvansky, 1998), they recognize simulation as an important component of language understanding. This chapter starts by describing the fundamental concept of mental simulation, namely, the Perceptual symbol systems (Barsalou, 1999), in section 2.1. Research on simulation’s necessity and function is essential; however, it is not the main concentration of the current study. Rather, simulation’s properties and plausible causes of match facilitations versus match inhibitions when simulation is involved in the critical responses will constitute this chapter’s primary focus. By utilizing fundamental research on the kinds of perceptual and motor properties that are internally simulated in language comprehenders’ mental representations of described events, simulation mechanisms are discussed in terms of the
temporal and spatial overlap between simulation and visual perception, including the ways in which these two critical factors interact with each other and influence the simulation process. The relative timing of the simulations and sensory-motor properties are crucial to the current study because it examines motor effects on simulation at different points in the time course of sentence production.

2.1 Perceptual symbol systems

Barsalou’s Perceptual symbol systems (1999), among other theories of Grounded Cognition, claim that we understand language by mentally simulating multimodal experiential knowledge including real-world perceptual and motor experiences, and that such simulation involves activation of our sensorimotor systems (Barsalou, 2008; Gallese, 2007; Lakoff, 1987; MacWhinney, 1999; Yeh & Barsalou, 2006). In order to understand language, we access concepts by running a simulation of perceptual and motor representations and incorporate the associated embodied sensorimotor information, thereby allowing us to experience and filter the received message more realistically. For example, when people read a novel, they may project themselves into particular characters in the story, seeing the described world through different perspectives and experiencing the emotional and physical conditions of those different characters (Bergen & Chang, 2005; Borghi, Glenberg, & Kaschak, 2004; Brunyé, Ditman, Mahoney, Augustyn, & Taylor, 2009). Creating internal imagery of the context may bring listeners to a deeper understanding of the writer’s intended meaning. This internal reactivation is possible because our cognitive mechanism is developed and grounded in bodily interactions with the environment; that is, our conceptual knowledge is acquired through
sensorimotor interactions with the world (Barsalou, 1999; Lakoff & Johnson, 1999). Therefore, we understand language because it evokes our previous perceptual and motor experiences with scenes similar to those it describes; in other words, we mentally simulate the perceptions and motor actions described in the words we hear or read. Perceptual symbol systems assume that reactivating the stored neural patterns enables comprehenders to create, perform, and manipulate perceptual representations.

Barsalou (1999) proposed three fundamental characteristics of a simulation mechanism. First, because words are acquired along with psychological or physical experiences, processing words automatically activates simulation by accessing the stored schematic experiences in the brain. Schematic experiences, i.e., similar experiences that are combined and treated together by ignoring the subtle differences among them, can be modified to create novel perceptual and motor simulations in understanding things one has never before experienced.

Second, activating knowledge about one stored schematic experience will also activate other experiences with which it has become associated in the comprehender’s mind (Yeh & Barsalou, 2006). For example, when accessing stored schematic knowledge about playing the piano, a person might simultaneously recall a particular piano-playing experience in which he also smelled an apple pie that had just come out of the oven, or overheard his neighbors laughing, or saw a beautiful bird resting on the window. These bits of information are irrelevant to the core knowledge acquired through playing the piano (e.g., how to read notes, move fingers, and control the loudness of the sound), but are still closely tied with individual piano experiences. Through the process of understanding the word *piano*, language comprehenders unconsciously access the holistic
properties surrounding this concept, including piano-related activities, a visual representation of a piano (e.g., color and shape), sensory-motor (e.g., auditory and tactile) knowledge, and associated background information like that described above (Myung, Blumstein, & Sedivy, 2006; Pecher, Zeelenberg, & Barsalou, 2003; Pecher, Zeelenberg, & Raaijmakers, 1998; Yee & Sedivy, 2006; Yeh & Barsalou, 2006). Such background information is an inherent part of our experiences; hence, comprehenders include background information in the rich multimodal simulations (e.g., smell, sound, or visual perception) that they create of the concepts they are processing. Therefore, mental representations reflect not only the accumulated knowledge of the concepts but also how language comprehenders interact with the world (Vigliocco & Vinson, 2005; Zwaan, 1999; Zwaan & Radvansky, 1998).

Third, simulations are dynamic conceptual processes because mental representations are flexible depending on contextual meaning. In other words, simulation is not a product of perceptual or motor imagery encoded into an individual word in a fixed manner, but rather a dynamic product of imagery induced by sentential meaning integration. The knowledge that is stored in people’s minds is flexibly integrated and the appropriate imagery is dynamically computed as more information becomes available (Ferretti, McRae, & Kutas, 2007). For example, the sentence The ranger saw an eagle in the sky (e.g., a flying eagle) induces a different image than The ranger saw an eagle in its nest (e.g., a resting eagle) (Zwaan et al., 2002). Mental simulation is also dynamically performed by accessing inherent and useful information that has been gained through experiences. For example, the dynamic meaning construction in understanding a novel sentence such as Mary kicked the calculator involves not only the mediated action
depicted by the sentence (i.e., kicking a calculator) but also functional information inherent in the object (i.e., punching in a number) (Masson, Bub, & Warren, 2008). This is because the intrinsic object properties (i.e., the size and weight of a calculator, knowledge gained through previous hand-related experience with these objects) that are evoked by action representations are used as the basis for understanding what kicking the calculator would feel like. In other words, we understand and internally simulate novel events by retrieving object properties acquired through direct experiences.

The following sections summarize behavioral evidence as well as neural imaging evidence that empirically supports the Perceptual symbol systems.

2.2 Perceptual simulation of non-motion features in comprehension

Recent studies of mental representation have consistently shown that meaning comprehension heavily relies on perceptually simulating the information encoded in language and have yielded various insights into the ways in which mental simulation plays a part in language comprehension. Studies conducted by Zwaan and colleagues are striking because they do not investigate whether participants can match words with pictures, but rather if they can integrate meanings of separate phrases into a mental representation of the object’s implied attributes. In a series of picture verification experiments, where participants read a sentence and decided whether or not the pictured object was previously mentioned, the authors found that response time was faster when images matched the sentence’s implied orientation (e.g., *He hammered the nail into the floor*, for vertical orientation vs. *He hammered the nail into the wall*, for horizontal orientation; Stanfield & Zwaan, 2001), shape (e.g., *The ranger saw the eagle in the sky*,
for an eagle with wings outstretched vs. The ranger saw the eagle in its nest, for an eagle with wings folded; Zwaan et al., 2002), or visual resolution (e.g., Through the fogged goggles, the skier could hardly identify the moose, for a low resolution image vs. Through the clear goggles, the skier could easily identify the moose, for a high resolution image; Yaxley & Zwaan, 2007). These studies suggest that an object’s orientation, shape, and visual resolution are essential constituents of the mental representations that are generated during language comprehension.

Previous studies on mental simulation have focused heavily on the question of what kinds of perceptual properties are represented in mental images, and to what degree of detail, during sentence comprehension. They do not explore the processing mechanism involved in mental image construction. Although these studies do not explicitly describe the simulation mechanisms, one of the dominant interpretations regarding the construction processes is that language comprehenders formulate a single mental representation after deep semantic integration of different linguistic components. However, Connell and Lynott (2007) claim that this is not the case and that language comprehenders can simultaneously represent prototypical and atypical meaning-based imagery after they have heard a complete sentence. They investigate how comprehenders’ knowledge of the canonical color of certain objects interacts with contextual color information during language comprehension, and raise two questions.

The first question concerns the lexical level: when an individual word such as bear is independently processed, do comprehenders only evoke the perceptual properties of a canonical or typical bear (that is, a big brown bear), or do they evoke multiple types of bears including white bears or teddy bears? The second question is at the sentential
level: how is a perceptual representation constructed when a word such as bear is embedded in a sentence such as A bear was found in the North Pole? Does the representation change from a typical brown bear to a white bear because the North Pole suggests this color, or are representations of both the canonical and context-specific bears maintained in a parallel manner?

A modified version of the semantic Stroop paradigm has been used to test whether both typical and context-specific color representations are maintained at the sentential end. Participants read a sentence evoking either a noncanonical green color (Jane tasted the tomato before it was ready to eat) or a canonical red color (Jane tasted the tomato when it was ready to eat). Then they are asked to name the color of the word tomato by using a simple color term (e.g., red, green). In naming time, no effect of color implied in the sentences is observed for the typical red color, but the effect is observed for the atypical green color. That is, a context-color match facilitation is observed only when the color is atypical (e.g., green for a tomato), but no difference is observed when it is typical (e.g., red for a tomato). These results reveal that our conceptual knowledge of an object’s typical color is automatically activated by accessing an individual lexical item, and that this knowledge remains in our minds unconditionally, regardless of context. Atypical context-specific green color is selectively represented, if there is a biasing context that induces atypical knowledge, and it can be represented in parallel with a typical color representation. In other words, contexts do not override the intrinsic typical color of an object; rather, they coactivate multiple color representations. Maintenance of multiple representations is not specific to simulation mechanisms. This is also observed in other processing mechanisms such as semantic interpretations; that is, an
incorrect interpretation due to the initial understanding of a garden path sentence persists while the global correct interpretation is simultaneously computed (Christianson, Hollingworth, Halliwell, & Ferreira, 2001; Sturt, 2007).

One thing to note is that maintaining both representations is aided by the fact that they are identical except for their color (e.g., red vs. green). A color property, which does not contribute to the object configuration, is considered less prominent and is easily ignored when other, more stable properties such as an object’s shape are attended to or focused on (Proverbio, Burco, del Zotto, & Zani, 2004); thus, simultaneous representations with different colors may become possible. The notion that the property of color has distinct characteristics from the content properties that influence objects’ configurations, such as shape, gains empirical support from behavioral and event related potential (ERP) experiments conducted by Proverbio, Burco, del Zotto, and Zani (2004). They show that color discrimination is processed faster, and thus more easily, than shape discrimination, but that color variations have no effect on the categorization process. Thus, shape recognition does not rely on color features, that is to say, intrinsic color properties are not central constituents of objects. Therefore, it may be premature to conclude that parallel representations are generally supported within perceptual simulations before we explore whether multiple representations that differ in stable, central properties are also simultaneously maintained during.

Additional converging evidence for perceptual simulation is provided by Borghi, Glenberg, and Kaschak (2004), who demonstrate that contextual information flexibly changes comprehenders’ perspectives in simulating situated motor actions of described events. In a part-verification test, participants who read a sentence designed to evoke
either an inside perspective (e.g., You are driving a car) or an outside perspective (e.g., You are washing a car) decided whether the subsequent word (e.g., steering wheel) was a part of the object in question (e.g., a car). The results show that car-interior words (e.g., a steering wheel, a horn) were identified faster than outside-part words (e.g., a tank, a tire) when an inside perspective of “driving a car” was taken, and vice versa. Interestingly, within a single perspective (e.g., a fueling perspective), participants respond faster to an object that is physically closer (e.g., a tank) than to one at a greater distance (e.g., tires).

The fact that a wide range of methodologies, including a Stroop task (Connell & Lynott, 2007), a picture naming task (Zwaan et al., 2002), and a part-verification test (Borghi et al., 2004) consistently indicate the presence of mental simulation implies that these results are not dependent upon any particular task. The primary tool used in perceptual simulation research—a picture verification task developed by Zwaan and colleagues—is commonly criticized. Detractors claim that its participants learn a strategy: they generate a perceptual representation after listening to the given sentence and comparing it with the provided picture. They adopt this strategy because a picture is routinely given after each sentence. If the results shown in the previous studies could be simply due to this image-generation strategy, then this would indicate that simulations are not automatic or necessary components of sentence comprehension.

However, there are two points against this argument. First, it is questionable whether participants generate images of the described scenes, as doing so will not help them respond to half of the questions (i.e., fillers) they receive. Second, although making a judgment about whether or not the pictured object is mentioned in the prior sentence does not require participants to form a representation of an object’s specific shape or
orientation, the detailed perceptual information does play an important role in recognition time. This indicates that detailed perceptual information is automatically activated during sentence comprehension in general, not only in response to some particular experimental task, even though doing so does not facilitate answering the question (see Connell & Lynott (2007) for an argument; Stanfield & Zwaan, 2001; Zwaan et al., 2002; Zwaan, Madden, Yaxley, & Aveyard, 2004).

2.3 Perceptual simulation of motion features in comprehension

Language enables comprehenders to perform not only perceptual simulations of non-motion features such as object properties, but also perceptual simulation of motor features involving physical interactions with objects. Neural imaging studies show that the premotor cortex is internally organized and specialized for specific motor actions such as hand-, foot-, and mouth-related actions (Pulvermüller et al., 2001; Tettamanti et al., 2005). Ehrsson, Geyer, and Naito (2003) use functional magnetic resonance imaging (fMRI) to demonstrate that both perceiving an action and simulating motor imagery of finger-, toe-, and tongue-specific movements systematically recruit the same spatial section of the brain that is used during execution of corresponding movements.

These studies, which have shown that understanding motor language involves neural activation similar to the corresponding sensorimotor action, further strengthen the claim that motor language comprehension is grounded in physical motor actions (Barsalou, 1999; Bergen, Narayan, & Feldman, 2003; Zwaan, 2004).

The theory of Perceptual symbol systems first suggested the tight link between language and experience. This link suggests the hypotheses that (a) we store and
generalize our perceptual and motor experiences during our interactions with the world, and (b) this stored knowledge encoded in our neural network is associated with words in our daily life, and (c) comprehending linguistic materials reactivates the stored experiential knowledge, resulting in creation of a perceptual or motor simulation of a described scene (Glenberg & Kaschak, 2002; Glenberg & Robertson, 2000; Zwaan et al., 2004).

The first part of this section focuses on the behavioral studies of perceptual simulation of motion features, paying close attention to causes of facilitative and inhibitory effects of simulations in sentence comprehension. Then, the second section discusses what non-linguistic information is dynamically incorporated in language comprehension and how. The last section presents some findings that show motor simulation, specifically focusing on the Action Compatibility Effect (ACE) in sentence comprehension and defining its critical role in the sentence production-based research introduced in Chapters 4–6.

2.3.1 Timing aspects of conceptual processing and sensorimotor processing

In order to investigate the presence of mental simulation with respect to linguistic materials, researchers have dominantly used both auditory/visual linguistic materials to induce mental simulation (i.e., conceptual processing) and either pictorial stimuli to investigate the details of perceptual simulation including object properties (Stanfield & Zwaan, 2001; Yaxley & Zwaan, 2007; Zwaan et al., 2002) or lexical items (e.g., a single word) to investigate the appropriate motions within the particular spatial representations (Borghi, Glenberg, & Kaschak, 2004; Zwaan, Madden, Yaxley, & Aveyard, 2004). One
of the most intriguing aspects of utilizing both conceptual and visual/lexical components is how to manipulate the relative time at which the two stimuli become available to comprehenders.

Use of this timing aspect in experiments can provide further support for the fundamental claim that mental simulation is evoked by recruiting neural activation similar to the corresponding sensorimotor action. If perceptual simulation is cognitively real so that it involves neural activation patterns similar to the actual sensorimotor (e.g., perceptual) processes during sentence comprehension, then subsequent perceptual processes should be easy and fast. This is because particular neural patterns that perceptual simulation previously utilized remain conceptually salient and accessible when language comprehenders process compatible visual information that involves the reactivation of the previously activated neural resources (Zwaan et al., 2004).

In contrast, if perceptual simulation and perceptual processes are simultaneously accessing and recruiting similar neural resources, then the simultaneous accessing of the single resource should be difficult and so will hinder the perceptual processes that are critical to responding in a sensible judgment task (Kaschak et al., 2005). Further explanations will be given below.

2.3.2 Language-mediated perceptual simulation of motion features

Match “facilitation” in sequential processing of simulation and visual stimuli

When language-mediated mental representations and visual percepts are sequentially processed, a match advantage is expected. That is, sentence processing becomes easier when the visual perception of motion matches the described motor events
(Zwaan et al., 2004). The match advantage is observed because motor simulations evoked by sentences create activation in a neural network, resulting in a processing facilitation of the subsequent compatible visual percept of motor events that involves a reactivation of the same neural pattern. In other words, previous activations leave residual potential that allows the same pattern to be reactivated easily.

For example, Zwaan, Madden, Yaxley, and Aveyard (2004) showed that sentences describing a toward-motion (e.g., The shortstop hurled the softball at you) or away-motion (e.g., You hurled the softball at the shortstop) facilitate the perception of sequential pictures depicting the corresponding object motion (e.g., a ball approaching vs. a ball going away). After listening to a sentence, a sequence of two pictures is displayed on the screen. A toward-motion is depicted by a small object followed by a medium one, while a large object followed by a medium one represents an away-motion. Participants determine whether or not the two pictures are of the same object; thus the expected response for all critical items is “yes.” The researchers found that participants’ picture-comparison judgment was faster when the linguistically described motion and the depicted motion matched than it was when they mismatched. This indicates that understanding described movements causes comprehenders to dynamically represent motor simulations.

While a reactivation of similar neural areas promotes the perceptual process of motor events, an activation of related but nonidentical neural activations hinders the subsequent semantic processes. For instance, in a lexical decision task, Bergen, Narayan, and Feldman (2003) used their behavioral results to argue that understanding motion verbs required activation of effector-specific neural structures. They found a mismatch
inhibition for a functionally related verb that was represented after a depiction of an event. Participants saw a line drawing picture followed by a verb and decided if the verb described the depicted event. When the picture and verb did not match, mismatch inhibition occurred. They postulated that different neural resources were simultaneously activated. The degree of inhibition differed depending on the effector involved in the action verb. More specifically, depicted events (e.g., scratch) and action verbs (e.g., hold) that shared the same effector (e.g., a hand) activated not exactly identical, but somewhat related neural resources, resulting in a stronger inhibition. Conversely, the researchers observed that events and verbs with different effectors (e.g., the event scratch with a hand effector and the event stumble with a foot effector) produced a weaker inhibition, presumably because they activated less closely associated neural structures.

*Match “inhibition” in simultaneous processing of simulation and visual stimuli*

Interference or match inhibition is expected in simultaneous processing of visual stimuli and a described scene that share the same motor actions. If Perceptual Symbol theory is true and language comprehension is grounded in sensorimotor systems, then the same neural activation patterns are engaged in both meaning comprehension (i.e., perceptual simulation evoked by sentences) and visual processing (i.e., perception of pictorial motor information)—an overlap that could result in interference if comprehenders need both simultaneously. For example, Kaschak, Madden, Therriault, Yaxley, Aveyard, Blanchard, and Zwaan (2005) explore how perceptual stimuli depicting one of four motions (i.e., horizontal bars moving up or down and spirals moving toward or away) interact with described motions including up (e.g., *The rocket blasted off*), down
(e.g., *The confetti fell on the parade*), toward (e.g., *The car approached you*), and away (e.g., *The car left you in the dust*) in generating a simulation. While participants are listening to sentences, they simultaneously see these moving pictures. Participants then decided whether sentences made sense (Experiment 1) and if they were grammatical (Experiment 2). The interference effect results indicate that the same system is engaged in processing both visual and internally simulated motion. This is congruent with the claim made by Perceptual Symbol theory, that is, the same neural activation patterns are engaged in visual processing (i.e., perception of pictorial motor information) and meaning comprehension (i.e., perceptual simulation evoked by sentences), and, hence, that language comprehension is grounded in sensorimotor systems. It is worth noting that this interference effect can also be explained by the non-integratability between linguistically motivated perceptual representations and visual percept. That is, when the contents of the sentence and of the percepts are not integratable, there should be an inhibitory effect because simultaneously processing two different contents—the percept and the simulation of the sentence—burdens the perceptual mechanisms (Kaschak et al., 2005; Richardson, Spivey, Barsalou, & McRae, 2003).

*Match “inhibition” in a visual field overlap between simulation and an unrelated object*

Another property evoked by perceptual simulation of motion properties is spatial information about the described event. Meanings of certain lexical items are closely associated with an iconic spatial relationship, for instance, the concept of *sky* is tightly coupled with *upness* while the concept of *ground* is tied to *downness*. This section
introduces two studies that instantiate that a spatial or directional overlap of linguistically motivated perceptual representations and the subsequent visual percepts cause an interference effect.

Both studies discussed here use a visual categorization task whereby participants are asked to read a sentence and then identify an unrelated object (e.g., a circle or a square). In Bergen et al.’s (2007) study, sentences containing either literal action verbs denoting dynamic upward/downward motions (e.g., climb, drop) or nouns denoting canonical up/down static locations (e.g., sky, ground) are presented followed by a picture of an unrelated object (i.e., a square or circle), which appears in one of four locations (top, bottom, left, or right). Match interference was observed when an object’s location matched the imagery of the corresponding sentence, e.g., objects at the top were identified more slowly after reading about an up-event than a down-event. This indicates that literal motion phrases can selectively simulate perceptual representations in corresponding locations of the visual field, and match inhibition occurs due to this spatial overlap. Similarly, Estes, Verges, and Barsalou (2008) demonstrated interference when a word (e.g., head, foot) that is displayed in the specific location that its meaning evokes (e.g., up for head, down for foot) is followed by an unrelated letter in the same location. This suggests that when a word (e.g., cowboy hat) forms a perceptual simulation in a particular location (e.g., up), then perceiving an unrelated letter (e.g., the letter ‘x’) that is subsequently represented in the same location causes interference effects due to the spatial overlap of processing two pieces of perceptual information (e.g., cowboy hat and x).
These comprehension studies introduced in this section highlight the fact that the timing of evoking simulations and perceiving given stimuli has considerable impact on the conceptual and perceptual processes (as reflected in facilitative or inhibitory effects). This is a key factor manipulated in the two series of experiments that this dissertation presents. That is, this dissertation investigates how physical motions influence simulation processes—and, ultimately, message and language production—by changing the timing of motor execution in the time course of message and language generation. This relationship between motion and simulation is examined in detail in Chapters 5 and 6.

2.4 Language-mediated motor simulation: Action Compatibility Effect (ACE)

This section introduces the concept of motor simulation and outlines how the internal experience of described events utilizes comprehenders’ inherent or learned knowledge of the mentioned objects or events. More specifically, it describes the action-sentence compatibility effect (ACE) proposed by Glenberg and Kaschak (2002), whereby comprehension of language stimulates conceptual knowledge of an appropriate action for a given situation and generates motor simulation. This ACE effect is the fundamental premise of my dissertation, which investigates the internal linkage between physical motion and linguistic processes, or more specifically, event generation processes (Chapter 5) and event description processes (Chapter 6) in sentence production. Furthermore, this section highlights one of the key factors manipulated in motor simulation studies, that is, the relative timing of hearing/reading linguistic materials and enacting actual motions, and explicates how the aspect of timing influences the facilitation effects in the meaning-motion match condition.
Glenberg and Kaschak (2002) employed a novel sentence verification experiment, where three buttons were arranged vertically on a keyboard, in order to examine the ACE effect. As soon as participants pressed down the middle button, a sentence appeared on a screen. Participants were asked to judge sentences as sensible or non-sensible by releasing the middle button and pressing either the button closer to their body or the one farther away. Half of the trials required away-movements for yes-responses while the other half required toward-movements for yes-responses. Match facilitation was observed when the implied toward- or away-movement simulation evoked by the sentences was compatible with the subsequent physical action (e.g., moving the hand toward or away from the body in order to press the response button).

The implication is that meaning is understood based on affordance, that is, possible interactions between actions and objects, so that understanding a toward-sentence, *Open the drawer*, results in the simulation of a pulling action toward the body, while reading an away-sentence, *Close the drawer*, evokes a pushing action in simulations. Since similar neural networks are responsible for executing the physical action and simulating the implied motion in a different time course, the residual activation due to consistency of the actions causes facilitation. As discussed earlier in this chapter, this match facilitation is observed when mental simulation evoked by linguistic input and visual information perceived in pictures are not activated simultaneously, but instead sequentially, in experimental tasks such as visual verifications or categorizations done after conceptual simulations were generated.

ACE effects have been observed in imperative sentences, concrete transfer sentences including double object (e.g., *You handed Courtney the notebook*) and dative
constructions (e.g., *Andy delivered the pizza to you*), and abstract transfer sentences including double object (e.g., *Liz told you the story*) and dative constructions (e.g., *The policeman radioed the message to you*). In short, Glenberg and Kaschak’s study shows that (1) understanding language describing motor actions generates motor simulation, which makes comprehenders re-experience the depicted event in their minds, and (2) motor simulation facilitates the subsequent physical action if motor simulation generated by comprehending sentences matches the subsequent physical enactment.

The motor simulation can be more detailed (e.g., an appropriate hand shape with a particular object) than general toward- or away-hand motion. For instance, Tucker and Ellis (1998, 2004) show that simply perceiving an object (e.g., a handle of a mug) mentally postulates potential actions with that object (e.g., grasping a mug with one’s hand). In Ellis and Tucker (2000), participants are instructed to remember the visual presentation of an object (e.g., a *grape*, which is intended to induce a simulation of a precision grip, or a *hammer*, which is meant to produce a simulation of a power grip) that remains on the screen for 700 ms. Then, participants hear either a high or low tone in which they categorize and indicate its auditory property by responding with a precision (i.e., a manual holding action with index finger and thumb) or power grip (i.e., a manual holding action involving palm and fingers). The results show match facilitations that are compatible with the match facilitation effects observed in perceptual simulation studies. In the same line of argument discussed in section 2.3.1, this match effect indicates that motor simulations activate particular neural networks and leave the conceptual traces for the subsequent neural re-activations that are prompted by actual motor enactments (e.g., precision or power grip).
Bergen and Wheeler (2005) designed their experiment so that each response involves a specific handshape that is made after a sentence has been comprehended. This study showed sentence-motion match effects that indicated that in addition to the general direction of arm movements (Glenberg & Kaschak, 2002), finer motor information, namely, the implied handshape (e.g., a flat handshape as described in *Paul carried the watermelon* or a fist as described in *Sue carried the marble*), is also simulated in understanding language.

Zwaan and Taylor (2006) extended the notion of the ACE effects to manual rotation. They utilized a knob task where participants manually turned a knob in either direction (i.e., clockwise or counterclockwise). In their experiment, participants heard a sentence that implied either clockwise motion (e.g., *Jane started the car*) or counterclockwise motion (e.g., *Julia set the clock back*). Subsequently, they made sensibility judgments about those manual rotation sentences by, for half the subjects, turning the knob to the right for a yes response and to the left for a no response, or, for the other half of the subjects, doing the reverse. If comprehenders internally experienced the meaning of the sentence while reading, then the residual activation generated by motor simulation during sentence comprehension should facilitate the knob rotation when the directionality between manual motion and the implied meaning matched, and in fact, this is what Zwaan and Taylor found. These results provide further evidence for the notion that simulation of an appropriate motion can be as fine grained as turning or rotating the hand in a specific direction.

One of Zwaan and Taylor’s (2006) experiments shows that this ACE match effect is also observed when reading and manual rotations are simultaneously performed. In this
experiment, participants manually rotate the knob counterclockwise or clockwise; each 5° of rotation causes the appearance of a new frame (i.e., 1–3 words) on the screen and the disappearance of the previous one (e.g., To quench/his/thirst/the/marathon/runner/eagerly/opened/the/water bottle, where each slash indicates a frame boundary). The results show that participants rotate their hand faster when physical rotation and direction are implied by sentence match. Critically, such match effect is observed at the verb region (i.e., opened) where the direction of the manual rotation becomes clear, but it quickly diminishes at the sentence final region (i.e., water bottle).

These results (i.e., the ACE effects observed in sensibility judgments conducted at the end of a sentence and in natural reading as soon as a verb specifies the directional information) suggest that language understanders experience motor simulation (a) in deciding the sensibility of the sentences because it involves a brief resimulation of the described motor contents at the end of the sentence, and (b) in incrementally unfolding the meaning of words.

2.5 Physical information influences simulation during sentence comprehension

Language processors seem to be sensitive to comprehenders’ physical situations, including their point of view (perspective) and their bodily status (e.g., posture or facial expressions), and they generate simulations by accommodating such non-linguistic information to some degree. For example, language comprehenders immediately employ their current perspective and reflect it in simulating the perspective of another person. Ramenzoni, Riley, Shockley, and Davis (2007) provide evidence that simulation of other
agents’ actions is formed via the observer’s own action capabilities and on-line perceptual information. In their experiment, an observer is asked to estimate another person’s maximum overhead reaching height in three different eye-height conditions: changes in observer’s eye-height by 0, 7.5, and 15 cm. The results show that an observer’s estimate of another person’s maximum reaching height increases as his own eye-height increases. In other words, manipulated eye-height information immediately influences the observer’s judgment of another person’s possible actions. This indicates that one predicts and understands another person’s possible actions by incorporating ongoing perceptual information and projecting one’s own current action capabilities onto the other. This is consistent with a simulation theory, which states that perceiving another agent’s performance activates the observer’s neural simulation of performing the same action. People are sensitive to a new perceptual environment and are able to use it immediately and dynamically in simulating another agent’s future possible actions.

More importantly, language processors are also influenced by their immediate bodily states (e.g., body postures, facial expressions associated with particular emotions), and this non-linguistic, physical information unconsciously elicits associated mental simulation, ultimately influencing the process of language comprehension (Barsalou, Niedenthal, Barbey, & Ruppert, 2003). Just as nodding one’s head produces positive emotional affect (Wells & Petty, 1980), holding a pen in one’s mouth in a way that forcefully creates a smile or frown evokes the associated emotion (i.e., pleasant vs. unpleasant), which in turn affects one’s speed in comprehending sentences describing pleasant or unpleasant scenes (Havas, Glenberg, & Rinck, 2007). These studies clearly show that bodily states are deeply tied to certain situations or emotions and crucially
illuminate causal effects of physical and mental experiences evoked by associated bodily states on the subsequent language processes.

Researchers have shown converging evidence that physical information is dynamically incorporated in language comprehension because language is acquired through physical experiences; thus, any bodily status that is tightly associated with language will have causal effects on simulation that, in combination with timing, influences linguistic processes. The causal role of these significant physical effects on subsequent language comprehension raises a critical question to be addressed by this dissertation: do body motions influence the subsequent conceptual process of unconstrained message construction (Chapter 5) and constrained message construction (Chapter 6) in language production? In fact, this dissertation proposes, and provides reasons to believe, that motor actions (which evoke motor simulation) may play a causal role not only in comprehending language, but also in producing language.
CHAPTER 3
MENTAL SIMULATION IN PRODUCTION

The first section of this chapter explains the reasons why there are limited simulation studies in language production compared to those in language comprehension. The second section presents research proposing gestures as an active manifestation of mental simulation during speech production. Then, various non-linguistic factors that influence the processes of putting/transferring thoughts into speech (e.g., word, structure, and word order) are outlined. Finally, the temporal order within mental simulation is proposed as an influential factor in determining word order in language production in Japanese, a language with relatively flexible word order.

3.1 Impediments inherent in research on simulation in production

Although the previous chapter discussed behavioral and neural imaging evidence indicating that simulation is a key component of language comprehension, recent studies investigating the role of simulation in formulating utterances are very limited in both their number and methodology. In fact, compared to the amount of information available from a vast number of simulation studies for comprehension, we know relatively little about perceptual or motor simulations in sentence production. This imbalance in the size and depth of simulation research in comprehension vs. production is due to the methodological difficulties inherent in studying mental simulation in language production. The existence of simulations during sentence comprehension can be tested by comparing response times when visual stimuli do and do not match prior sentence
descriptions (a picture verification task; see Chapter 2 for detailed description). Similarly, the existence of motor simulations is evidenced by differences in response times when subjects perform motor actions that are either compatible or incompatible with the implied motion of given sentences (e.g., the ACE effect proposed by Glenberg & Kaschak, 2002, described in Chapter 2). In both cases, it is clear that reading sentences is the cause of perceptual and motor simulations, and they are easily measurable by response time.

Three primary obstacles exist in the active investigation of simulation in sentence production. First, as opposed to simulation in sentence comprehension, we cannot use linguistic materials to generate mental simulations because the central question is whether speakers are engaged in simulation while they are freely transforming thoughts into language. Second, in order to ensure that the linguistic outcomes reflect the presence or effects of simulation, non-linguistic manipulation must be used as a prime source. However, it is difficult to ensure that non-linguistic manipulation successfully influences simulations, so that the manipulated simulation can be reflected in sentence production. Therefore, new non-linguistic tasks that definitely evoke simulations, but clearly do not activate associated linguistic elements must be created. A detailed description of such a new task, created for this study, is given in Chapter 4. Finally, it is difficult to analyze the free forms of production because this requires careful, objective evaluation of what the produced sentences really indicate about simulation. This concern about objectivity was met in the current research by establishing systematic criteria for sentence-type categorizations and employing multiple researchers who coded sentence types
independently; the methods used for categorizing and coding are explained in section 5.2.4.

The following section briefly outlines the widely accepted model of sentence production proposed by Bock and Levelt (1994) in order to show how a speaker’s intended non-linguistic message is transferred into linguistic form. It specifically focuses on the content of the intended message represented at the conceptual level because the current research investigates whether or not external factors (i.e., body motion) can unintentionally influence message formulation.

### 3.2 Sentence production model

Bock and Levelt (1994) provide a model that sketches how speakers produce an utterance. A diagram of this model is reproduced in Figure 3-1 below, adapted from Bock and Levelt, 1994, Ferreira and Slevc, 2007, and Levelt, 1999. Utterances are produced through three primary processes, including message formulation, grammatical encoding, and phonological encoding. Production begins when speakers formulate a prelinguistic concept of their intended message. Although variant models of sentence production exist in psycholinguistics, they seem to agree that there are three types of internal components that form a thought or message: perspective meaning, semantic meaning, and relational meaning (Bock, 1995; Bock & Levelt, 1994; Ferreira & Slevc, 2007; Levelt, 1999; Slobin, 1996).

First, “semantic meaning” represents the semantic knowledge of entities/participants in isolation, including each of their perceptual characteristics, statuses, or actions. That is, semantic features of each participant/entity are individually
represented, but their relation with respect to each other is not yet established. Second, “relational meaning” represents how the participants/entities in the event conceptually establish/configure the relationship among them. This may include a relational construction to determine who is performing what action on which entity in what situation; in other words, this conceptual process casts/assigns a relational role to each participant/entity in order to formulate a meaningful event. Finally, a message or event that is constituted by the relationship among participants/entities can be framed from different viewpoints. A single event can be perceived and experienced differently depending on which viewpoint or participant position one puts oneself in. This modulation of the viewpoint in the event is called “perspective meaning.” In sum, meanings are fully formulated by specifying each of three conceptual components, and these processes produce the conceptual event that represents the information of “who did what to whom.” It is worth noticing that the underlying assumption of this meaning construction is that the message emerges from speakers’ conscious intention and reflects their intended meaning.

These non-linguistic conceptual materials are delivered to the subsequent linguistic encoding processes, including grammatical encoding and phonological encoding. In grammatical encoding, speakers first select appropriate lexical entries, and then determine the syntactic structure that will frame and order these lexical entries. These two stages of grammatical encoding are called functional processing and positional processing, as shown in Figure 3-1. Functional processing consists of lexical selection, where speakers choose lexical entries that reflect their intended meaning and the lexical entries activate their enclosed grammatical and semantic information (i.e., lemmas), and
function assignment, which determines the grammatical roles (e.g., subject, object) of the lexical entries. Positional processing involves constituent assembly and inflection, where the selected lexical entities become associated with appropriate grammatical inflections (e.g., tense and aspect) and are arranged into a suitable order. Finally, speakers retrieve the phonological information of the grammatically encoded message in order to articulate their utterance.

Figure 3-1: Model of sentence production (adapted from Bock & Levelt, 1994; Ferreira & Slevc, 2007; Levelt, 1999)
3.3 Possible effect of external factors on thought and speech

Previous sentence production studies in psycholinguistics have proposed three types of information or internal components (i.e., perspective meaning, semantic meaning, and relational meaning) that form a thought (Bock, 1995; Bock & Levelt, 1994; Ferreira & Slevc, 2007), but none of the production models have addressed the question of what external factors might shape our thought. Existing models primarily regard speakers’ intended messages as emerging from their own intention, but they have not taken into account non-detachable or inevitable external factors, including speakers’ surrounding environment and the physical activities with which they are occupied. When people produce language in daily life, they are commonly engaged in some form of physical activity. They may speak while cutting vegetables, organizing books, or watching a football game. In addition to the speaker’s intended message, these unintended, unplanned non-linguistic factors (e.g., concurrent physical activities or perceptual or motor simulations caused by perceiving an event in the surrounding situation/discourse) may affect his ongoing cognitive status, shape the emerging simulation, and ultimately influence the speech outcome. In fact, previous comprehension studies have shown that physical engagement influences comprehenders’ interpretations of ambiguous sentences (Boroditsky & Ramscar, 2002) and their lexical selection of synonyms (Tseng, Hu, Han, & Bergen, 2007). Therefore, engagement in physical activity also seems to be a plausible means of activating simulation in sentence production. The current research focuses on how simple direction-specific hand motion evokes direction-
oriented simulation that readily configures or affects the relational event construction of the message and influences speakers’ linguistic choices.

3.4 Syntactic priming as a tool to illuminate the relationships between conceptual and grammatical encodings in the production process

As explained above and shown in Figure 3-1, a production process can be reasonably grouped into a sequence of three major steps, that is, conceptual, grammatical, and phonological encodings. However, there has long been debate over how to define how each step communicates with the others in producing language.

The phenomenon of syntactic priming can be used as a tool to investigate the intriguing question of whether conceptual and grammatical encodings have independent or interactive relationships. Syntactic priming occurs when a prior utterance that a speaker produces implicitly affects subsequent speech (Bock, 1986; Bock & Loebell, 1990). The repetition of the syntactic structure is observed when speakers have to choose a particular syntactic structure when alternative constructions are available to convey their intended message (e.g., *Lightening is striking a church* vs. *The church is being struck by lightening*; Bock & Warren, 1985). Because syntactic priming occurs at the syntactic level, which is directly connected to the conceptual level, investigating whether changes in conceptual features can change the impact of syntactic priming enables us to see whether the conceptual and syntactic representations are distinct or inseparable mechanisms in language production systems. More specifically, if some interactions take place between the two levels, then changes in conceptual components will result in changes in syntactic structure selections or changes in the size of the syntactic priming
effect. However, no such changes due to the conceptual factors are expected if syntactic representations are independent from conceptual factors.

Syntactic priming appears to be a quite general and pervasive process in our language mechanism because consistent priming effects have been reported with different grammatical constructions (e.g., double object and prepositional dative constructions, active and passive constructions), in different types of languages (e.g., head-initial languages including English and German, and head-final languages including Japanese), and in a wide range of methodologies including picture description (Bock, 1986; Bock & Warren, 1985), sentence recall (Ferreira & Yoshita, 2003), and sentence completion in written and spoken language (Pickering & Branigan, 1998; Yamashita & Chang, 2001).

In order to answer the question of whether syntactic representations can be completely isolated from conceptual factors, there are two contradictory accounts, that is, a form-mapping account and a meaning-mapping account. First, the form-mapping account (Bock & Loebell, 1990) claims that abstract syntactic structures (i.e., form) are separable from non-linguistic conception (i.e., meaning) and have no interactive relations in the process of generating language. Therefore, changes in the conceptual factors will not affect the subsequent process of syntactic construction. This account predicts that syntactic priming is simply the result of syntactic repetition, and that conceptual overlap between prime and target sentences does not affect syntactic priming.

For example, Bock and Loebell (1990) utilize a memory and sentence description task where, after perceiving a prime sentence, participants describe an event depicted in a picture that allows both prepositional dative and double object constructions. The results show equivalent priming between prime and (prepositional dative) target sentence pairs
even though the prime sentences involve different thematic roles (e.g., locative as in (a) and beneficiary as in (b)) and verb types (e.g., transitive vs. ditransitive):

a. The wealthy widow drove her Mercedes to the church. [transitive with locative to-phrase]
b. The wealthy widow gave her Mercedes to the church. [ditransitive with beneficiary to-phrase]

Similarly, although the conceptual structures of passive by-phrases such as (c) and locative by-phrases such as (d) are thematically different, both of them equally primed the consequent production of passive sentences:

c. The 747 was alerted by the airport’s control tower. [passive by-phrase]
d. The 747 was landing by the airport’s control tower. [locative by-phrase]

These results suggest that the differences in the thematic roles, which are one of the conceptual features, are not critical factors that affect syntactic priming.

However, syntactic priming is not simply due to surface configuration similarities, but rather to constituent structure similarities. For example, sentences with a prepositional dative as in (e) and infinitive structures as in (f) appear to be similar as to surface structures, but only prior exposure to prepositional dative sentences primes utterances with the prepositional dative structures; exposure to infinitive structures does not, due to their constituent structure differences:

e. Susan brought a book to Stella. [locative to-phrase]
f. Susan brought a book to study. [infinitive verb phrase]

To conclude, conceptual similarities (e.g., thematic roles) in prime-target pairs are not the basis for syntactic priming, while similarities in the hierarchal configurations at the grammatical level are the critical factor for syntactic priming effects. Because syntactic priming is insensitive to conceptual components but sensitive to grammatical
components, it confirms the form-mapping account that claims that the abstract syntactic representations are separable and independent from conceptual representations in language structure. A great number of syntactic priming studies have shown converging evidence that structural repetitions occur at the level of abstract syntactic structure, and the independent existence of abstract syntactic frames is generally agreed upon; however, there is an ongoing debate regarding how such abstract syntactic structure relates to other elements in the production system.

Second, the meaning-mapping account (Cleland & Pickering, 2003; Griffin & Weinstein-Tull, 2003) proposes that syntactic structures are directly linked to and communicative to the pre-linguistic message. Thus, certain changes in the non-linguistic properties that constitute the abstract message will have an effect on the process of grammatical encoding and change the syntactic structures and word order. This account assumes that conceptual components play a critical role in syntactic priming, so that conceptual overlap between prime-target pairs will result in an enhancement of syntactic priming. This account is supported by a study done by Cleland and Pickering (2003), who utilize noun phrase structures in syntactic alternations between an adjective + noun phrase (e.g., a red sheep) and a relativized noun phrase (e.g., a sheep that is red). They show that semantic similarities between head nouns in prime (e.g., goat) and target (e.g., sheep) cause a significant enhancement of syntactic repetitions. However, phonological similarities between prime and target (e.g., sheep and ship) do not influence the magnitude of syntactic priming. This indicates that the more similar the conceptual components in primes and targets are, the more likely the message is processed and encoded into the same syntactic structure.
In addition, Griffin and Weinstein-Tull (2003) use a sentence recall task and demonstrate that a subtle difference in the conceptual level, that is, the number of thematic roles encoded in an argument can affect syntactic priming. For example, one thematic role (i.e., an argument of ‘be nice’) is assigned to Mary in *John believed Mary to be nice*, while two thematic roles (i.e., an argument of ‘be nice’ as well as a persuadee) are assigned to Mary in *John persuaded Mary to be nice*.

The Isomorphic Mapping Hypothesis (O’Grady & Lee, 2005), which proposes a tight relationship between the chronological order of the event structure and word order (the next section provides an extensive description of this hypothesis), and its extended hypothesis regarding simulation–word order correspondences in this current study are conceptually compatible with the meaning-mapping account. This is because both hypotheses presuppose that conceptual and syntactic representations are not completely disconnected, but rather are interconnected; thus, conceptual factors (which are involved in the process of formulating speakers’ intended messages) influence how the message is grammatically encoded (e.g., word order).

### 3.5 Word order in simulation and non-linguistic factors that influence linguistic construction

This section introduces the Isomorphic Mapping Hypothesis (IMH) (O’Grady & Lee, 2005), which proposes a tight relationship between the chronological order of event structure and word order. In order to investigate whether the preference for a match between event language and word order is merely an off-line correlation or is in fact an on-line, causal effect of simulation, I examine the question of whether speakers’ embodied status that is manipulated by toward- or away-motor activities can influence
the choice of message formulation and event language (i.e., toward- or away-language) as well as corresponding word order (i.e., Loc-DO or DO-Loc word order). In addition, this section describes some non-linguistic factors that influence linguistic constructions because they inform the essential question of the current study, that is, whether physical motion, as a non-linguistic factor, can affect the message and linguistic encodings in sentence production.

O’Grady and Lee (2005) proposed the IMH after observing that English- and Korean-speaking participants with agrammatic aphasia (language impediments associated with lexical retrieval processes and with appropriate grammatical construction processes) are better at comprehending sentences in which the order of the linguistic arguments aligns with the way the corresponding event unfolds. In their act-out comprehension task, English-speaking Broca’s aphasic patients with agrammatism were instructed to respond to a locative pattern (e.g., *Put the crayon on the pencil* as in (a); examples in this section are adopted from O’Grady and Lee, 2005) whose theme-location pattern is isomorphic with the described event, or an instrumental construction (e.g., *Tap the crayon with the pencil* as in (b)) whose theme-instrument pattern does not follow isomorphic order.

(a) Locative pattern (isomorphic):

She put the crayon on the pencil.

| X acts on Y placing it on Z. | ← event |

| Agent | theme | location |

(b) Canonical instrumental pattern (non-isomorphic):
She tapped the crayon with the pencil.

\[
\text{X uses Y to tap Z.} \quad \leftarrow \text{event}
\]

As the IMH predicts, participants’ performance was better on the isomorphic pattern than on the non-isomorphic one.

Participants with agrammatism also performed better on a topicalized instrumental pattern (as in (c)) that manifests isomorphic order (e.g., *With the pencil, tap the crayon*) than on a canonical instrumental pattern that does not (e.g., *Tap the crayon with the pen* as in (b)), although the syntactic structure associated with the isomorphic pattern is less frequent and less basic compared to the canonical structure associated with the non-isomorphic pattern.

(c) Topicalized instrumental pattern (isomorphic):

\[
\text{With the pencil, tap the crayon.}
\]

\[
\text{X uses Y to tap Z.} \quad \leftarrow \text{event}
\]

O’Grady and Lee’s (2005) results are compatible with the results of a study by Cho, Lee, O’Grady, Song, Suzuki, and Yoshinaga (2002) that investigated the comprehension of Korean, whose flexible word order allows both the instrumental-accusative and accusative-instrumental orders are both acceptable like Japanese. Cho et al. show that Korean preschool children prefer the instrumental-accusative order in instrumental sentences in Korean (e.g., *Pen-Instrument pencil-Acc touch-Beneficial-
Future-Sentence ender in Korean ‘Will you touch the pencil with the pen?’) because the instrumental-accusative order is consistent with the order of how an agent interacts with the direct object and uses it to complete the described event. Crucially, however, they exhibit a preference for the reverse order, that is, the accusative-dative order in dative sentences (e.g., Handkerchief-Acc cap-Dative throw-beneficial-future-sentence ender in Korean, ‘Will you throw the handkerchief to the cap?’), although such an order is rarely produced in maternal speech to children. Japanese-speaking children exhibit a similar tendency with double object constructions in Japanese (Isobe et al., 2004).

O’Grady, Yamashita, and Lee (2005) further employ locative patterns (e.g., Put the crayon on the pencil), basic instrument patterns (e.g., Tap the crayon with the pencil), and fronted instrument patterns (e.g., With the pencil, tap the crayon) in a study that demonstrates that isomorphism is a critical factor in processing language. In their study, Korean and Japanese learners of English as a foreign language have a strong preference for the NP constructions that are temporally consistent with the event representations (including the locative pattern as well as the fronted instrument pattern, even though the fronted construction is infrequent) over the basic instrument pattern. Similar results have also been found with English-speaking learners of Japanese as a second language (Ito, 2007). These findings, drawn from effective manipulation of syntactic-construction frequency, can shed light on general processing mechanisms – that is, language users can process more easily if word order and the event representation are temporally aligned.

These studies suggest that isomorphic mappings between syntactic structure and event representation are responsible for word order preference. In other words, participants favor the word order that presents referents in an order consistent with the
corresponding situation described in the sentence. The principal concept that is rooted in
the IMH is fundamentally consistent with the proposal of Matsuoka (2003) and
Miyagawa and Tsujioka (2004) that the double object construction in Japanese has no
single canonical word order, although traditionally the goal-theme order has been the
basic word order while the theme-goal order has been considered a scrambled form
derived from the basic order (Hoji, 1985, cited in Isobe et al., 2004; Miyagawa &
refer to the underlying motivations for the general preference for a match between event
type and word order (while the IMH does), they descriptively state that there are two
distinct types of basic word order that are determined by the types of ditransitive verbs in
Japanese, namely, show-type verbs that have the sequence of dative-accusative argument
as a basic word order and pass-type verbs that have the sequence of accusative-dative
argument as the basic order.

Deriving the idea from the IMH (O’Grady & Lee, 2005), and in a similar line of
argument, this dissertation proposes that simulation that activates accumulated
sensorimotor experiential knowledge (e.g., physical motions) is the source of word order
preference. In other words, the preference for a match between event language or
particular types of verbs (e.g., show-type verbs vs. pass-type verbs) and word order is not
merely a correlation. Such a preference may provide evidence that simulation is an active
source for the modulation of word order in sentence production if the
preference/correlation ratio is boosted by physical motion that speakers are engaged in
while constructing the message. This indicates that speakers favor a word order that is
consistent with the construction of the real event because they are mentally engaged in re-experiencing the event while formulating the message and producing language.

When speakers construct a message and perform simulations in which speakers internally re-experience the described scene, the temporal knowledge of actions (that has been acquired through experience) will influence the arrangement of phrases (i.e., word order). For example, messages involving movements toward the speaker’s body (e.g., *I grab the milk from the refrigerator*) should require the speaker (during simulation) to first pay closer attention to the object denoted by the locative phrase (e.g., the refrigerator), then engage that object (e.g., open the refrigerator), and then contact the movable direct object (e.g., grab the milk). In the flexible word order of Japanese, this sequence should result in the production of more canonical sentences that reflect the corresponding flow of the event (i.e., Subject-Locative-Direct Object-Verb). Conversely, more non-canonical sentences (i.e., Subject-Direct Object-Locative-Verb) are expected in Japanese when speakers are expressing an event associated with movement away from the speaker’s body (e.g., *I put the milk in the refrigerator*) since this word order matches the conceptual order of the event. Since simulation is mentally performed by activating physical experiences (e.g., toward or away hand motions in order to complete the activity) in describing the event, physical actions can be integrated and used as a part of the simulation process. If motion boosts the use of a particular word order that corresponds to the temporal sequence of the event construction, then this supports the proposal that simulation, which involves a cognitive activation of physical experience, is the plausible reason behind the repeatedly observed correlation between event type and word order.

It is worthwhile to note that toward- or away-language does not necessarily
employ the word order that corresponds to the sequences of an event. For instance, one of the dominant properties of away-language is that an object is transferred to a particular destination that is often assumed to be in a distal location from the agent’s body (e.g., *putting the plate into the dishwasher*). However, an object can be transferred toward one’s body, instead of away from one’s body as indicated by the phrase “*placing the keys in my pocket.*” Likewise, utterances considered as toward-language tend to share the characteristics of transferring an object from a distal location toward one’s body (e.g., *getting the book from the bookshelf*). This is, however, not the only possibility; for example, the action implied by the phrase “*taking the pen from the penholder*” could be away-motion if one holds the penholder close to one’s body and takes a pen out of it.

Therefore, based on the norming study described in chapter 5, experimental stimuli are selectively set up so that “location” is consistently positioned physically away from the agent. This ensures that transferring an object from the location systematically produces toward-motion while moving an object to the location unquestionably produces away-motion.

The inclusion of Japanese as well as English allows these experiments to investigate how mental simulations during message formulation are reflected through the flexible word order of Japanese (i.e., word order can be used as an indicator of the event structures of mental simulation), and whether a natural order of the simulated event emerges that would not be readily seen in English, due to its fairly rigid word order. In other words, it is assumed that word order can be driven by the temporal order of a simulated event such as whether a certain entity has been mentally focused first in pursuing the described event.
In addition to word order determination, the current study utilizes non-linguistic physical activities to investigate their effects on meaning and linguistic constructions. The underlying assumption/hypothesis in the current study is that non-linguistic motor actions influence language construction processes. This assumption is based on some studies that show that non-linguistic information, such as the timing of visual information retrieval (Brown-Schmidt & Tanenhaus, 2006; Gleitman, January, Nappa, & Trueswell, 2007), highlights a certain aspect of a message during language formulation, thereby influencing linguistic encoding processes (e.g., structure and word order).

Brown-Schmidt and Tanenhaus (2006) claim that initial eye-landing site (or the order of retrieving visual information) influences message formulation and language use. This claim is evidenced by a simple language game in which speakers describe the target shape to their interlocutors and speakers’ eye movements are monitored. Eye movements indicate that when speakers recognize contrastive information early, they are likely to produce pronominal phrases with size adjectives such as the small horse. On the other hand, when speakers do not notice contrastive information before utterance onset, then they place size adjectives in a later position, leading to post-noun repairs such as the horse, uh small one.

Moreover, Gleitman, January, Nappa, and Trueswell (2007) propose that the initial visual attention influences the order of language encoding. They use an attention-capture manipulation (i.e., a brief flash that unconsciously directs the speaker’s initial eye-landing site to a particular part of the picture stimuli) to study whether manipulations of visual attention can influence sentence formulation when people describe a simple scene. The results show that the attention-capture manipulation affects the word order of
equally salient participants (e.g., *The dog and cat* … vs. *The cat and dog* …), the choice of active/passive descriptions for a single event (e.g., *The cat drinks the milk* vs. *The milk is drunk by the cat*), and the perspective selection of the event (e.g., *The dog is chasing the man* vs. *The man is running away/fleeing from the dog*), and other example verbs that describe a single event with different perspectives including *buy* vs. *sell*, *win* vs. *lose*, and *give* vs. *receive*). Also, speakers tend to encode the initially activated information as a subject in utterances describing scenes.

These studies indicate that the language processor and linguistic planning are sensitive to the timing of visual information retrieval and to the immediate or currently available information. In other words, the time at which non-linguistic information is retrieved (e.g., the time of visual apprehension) and the time of linguistic formulation are tightly coupled. Such findings that provide evidence for the susceptible or perceptive nature of language processors are crucial to this dissertation for two reasons.

First, if message planning is only sensitive to the order of receiving information from picture presentations, the retrieval order systematically determines the word order in utterances. For example, if speakers receive a location (e.g., a basket) followed by an object (e.g., an apple), then the order mentioned in the utterance should be the location-object that is identical to the perceived objects’ order. Crucially, however, if language processors are also sensitive to non-linguistic motion, then this should influence the event formulation whose isomorphic pattern eventually modulates word order, regardless of the fixed sequence of picture presentations.

Second, the spontaneous and effortless integration of non-linguistic information into message construction shown in previous studies suggests a tight temporal association
between non-linguistic and linguistic processes. If speakers are able to accommodate and immediately reflect non-linguistic information such as the initial visual attention in their utterances, then speakers should also be able to accommodate and integrate another type of non-linguistic information, namely, physical motion, in their message planning, and ultimately in their utterances.

The use of cross-linguistic experiments can indicate whether or not the role of motor action in message construction is a language-independent cognitive feature, depending upon whether or not the same results are observed across typologically differing languages. Moreover, revealing the interrelationship between body manipulation and language production (i.e., message types and word order) will shed light on the influence of motor actions on the human cognitive system, and, more generally, on the relationship between linguistic and non-linguistic cognition.

3.6 Gestures as an indicator of simulation during speech production

Within the limited number of studies on simulation in sentence production, gestures employed during speech have been used as the dominant (if not sole) indicator of active use of simulation during speech production. Section 3.3.1 first briefly summarizes the functional role of gestures in speech and discusses previous gesture/speech studies that propose that gestures emerge automatically as a byproduct of mental imagery during language production. Then, the Gestures as Simulated Action (GSA) framework, which illustrates the mechanisms of how gestures emerge during simulation in language production (Hostetter & Alibali, 2008), is discussed in section
3.3.2. This section ends with some plausible limitations underlying gesture-in-speech studies.

3.6.1 Functional role of gestures in speech

In the field of gesture research, there have been long and controversial arguments regarding the purpose of gestures during speech production. Some investigators claim that gestures are produced to enhance listeners’ understanding, i.e., for a communicative purpose between a speaker and a listener (Alibali, Flevares, & Goldin-Meadow, 1997). If this is the case, gesturing should decrease when a speaker cannot see the listener. Other researchers argue that gestures are performed in order to facilitate smooth speech generation, i.e., for the speakers’ internal purposes. If this is the case, visibility between a speaker and a listener should not influence the amount of gesturing during speech.

However, a recent study conducted by Alibali, Heath, and Myer (2001) empirically showed that the underlying purposes of gestures are for both communication and internal speech production. First, the fact that speakers produced fewer gestures when a screen limited visibility between them and their listeners indicates that gestures are generated with communicative intention. Second, however, gestures also seem to have a facilitative function for speech production, since (a) the decreased amount of gesturing increased the amount of disfluency, and (b) speakers still used some gestures even when they could not see their listener. In addition, the increased disfluency due to fewer gestures may have originated from the limited or less visual feedback from the listeners such as signs of comprehending (e.g., nodding, eye contact, or chiming in). Speakers may utilize such
visual feedback from listeners to facilitate their semantic and lexical access when articulating their message.

The goal of my dissertation, however, is not to identify the functional purpose of simulation during speech. Rather, I aim to explore the role of physical motion in simulations, and how these simulations affect the formulation of messages. Therefore, it is relevant to introduce the Gestures as Simulated Action (GSA) framework, which solely focuses on the simulation mechanism that facilitates the emergence of gestures, but not on the functional aspect of gestures (Hostetter & Alibali, 2008).

3.6.2 Gestures as Simulated Action (GSA)

This section introduces the Gestures as Simulated Action (GSA) framework (Hostetter & Alibali, 2008), which attempts to explain how gestures emerge during mental imagery and language production. The central claim of the Gestures as Simulated Action framework is that gestures emerge from simulations (i.e., perceptual and motor simulations) that are generated during language production. The underlying presupposition is that both thinking and speaking are rooted in the embodied cognitive system that generates mental imagery. Gestures result from the speaker’s active engagement in generating simulations during language production.

The GSA framework declares that there are three factors—neural, cognitive, and motor—that determine whether simulated concepts or events can be expressed/conveyed as gestures. First, as neural imaging studies in sentence comprehension have shown (see sections 2.2 and 2.3), generating simulations activates an appropriate area of premotor cortex, which is normally responsible for action planning and preparation. GSA suggests
that if the neural activation in the designated premotor area is strong enough, then this activation can potentially spread from premotor to motor areas, resulting in explicit physical motions (i.e., gestures). However, this spreading process cannot be accounted for solely by the strength of premotor activations or simulations, because it can vary depending on the strength of connections between premotor and motor areas. For example, if these two areas are tightly connected, then even weak activations in the premotor areas can be smoothly transmitted to the motor areas, and thus gestures are produced. On the other hand, if the connection is weak, even strong activations in premotor areas will not be realized as overt gestures. Second, the height of the speaker’s “gesture threshold” also determines the production of gestures. Speakers are able to adjust their gesture threshold depending on situations. Gestures emerge when the premotor activations are strong enough to exceed the boundary of their current gesture threshold. Third, since both articulatory planning and simulations involve premotor activations, once speech is articulated via spreading activations from premotor to motor areas, other concurrent premotor activations (i.e., simulations) also spread to motor areas. This may also be responsible for why people perform gestures during speech.

Some studies empirically support the GSA theory by claiming that two types of gestures, character-viewpoint (i.e., gesture produced by the speaker as if he were the character in the story) vs. observer-viewpoint (i.e., gesture produced by the speaker as if he were watching the character from a third person perspective) emerge from motor simulations and perceptual simulations, respectively (McNeill, 1992; Hostetter & Alibali, 2008). As one of the few pieces of empirical evidence for simulations in sentence production, a study by Parrill, Bullen, and Hoburg (2009) describes speakers’ use of...
character-viewpoint gestures in retelling a remembered story after viewing video clips or texts. For example, after watching a cartoon skunk in a video clip, a speaker describes the video to a listener by imitating the skunk. These character-viewpoint gestures could be evidence of simulations, but the possibility that they are simple re-creations of what is shown in the video clip cannot be overlooked. Therefore, comparable text descriptions of the video clips were created and the gestures the readers used in describing the readings to their listeners were compared to those produced after viewing the video clip. The authors propose that the fact that similar proportions of character-viewpoint gestures are found in response to video clips and texts suggests that gestures generated while describing a story are not re-creations; rather, they reflect the perceptual and motor representations created when producing language.

Casasanto and Lozano (2006) explore spontaneous gestures performed when retelling remembered stories involving literal spatial language (e.g., The rocket went higher...), metaphorical spatial language (e.g., The temperature went higher...), or non-spatial language (e.g., The temperature got hotter...). Gestural representations indicate what kind of conceptual simulations are activated while using spatial language, and gestural use indicates whether or not all three types of spatial language actively use the same concrete spatial domain of knowledge. The results show that consistent gestures (e.g., upward gestures) were employed when talking about literal and metaphorical spatial language, as well as non-spatial language. Forms of gestures reveal that abstract spatial language is grounded in more concrete concepts, namely experience-based motor action. This indicates that all three types of language are grounded in the same concrete representations. The fact that accompanying gestures were observed with non-spatial
language (e.g., *hot*) reveals that gestures are not lexically, but rather conceptually, motivated. The study suggests that regardless of whether or not explicit spatial language is used, speakers form experientially-based motor representations that are used to think and talk about spatial concepts.

The following section discusses a limitation of the gesture/speech studies that attempt to provide evidence for the active use of mental simulation in sentence production that have been described in this section.

### 3.6.3 Limitations underlying gesture-in-speech studies

Although the GSA framework and the empirical studies presented in section 3.3.2 suggest that gestures are a reflection of simulation in concurrent speech, they do not inform us whether or not gestures are generated as a part of the production process or as a supplement aimed at smooth utterance formulation. If gestures enhance the process of internal speech, then they function as a fuel for speech production or they work as glue to link the message and speech (or lexical encoding). In fact, the Lexical Access hypothesis (Krauss, Chen, & Chawla, 1996; Krauss, Chen, & Gottesman, 2000) suggests that gestures function as a bridge that links spatial- and motor-related thought and the retrieval of appropriate lexical items. This theory gains empirical support from the fact that more gestures are employed when speakers have difficulty in accessing the appropriate lexeme (Morsella & Krauss, 2004) and that preventing speakers from gesturing results in a higher rate of disfluencies (Rauscher, Krauss, & Chen, 1996). Although the Lexical Access hypothesis proposes that simulation is performed to assist the appropriate linguistic encoding process for the intended message and serves to
facilitate smooth utterance formulation, it still does not eliminate the possibility that gestures are one of the essential components in the sentence production system for formulating speech. That is, simulation is generated because it is one of the components that are embedded in the sentence production mechanism.

In order to have a clear picture of how simulations fit into the production mechanism, an underlying methodological issue in investigating the link between gestures and speech should be addressed. Parrill, Bullen, and Hoburg (2009) proposes that gestures are the result of simulations underlying language production. Although their task in fact involves language production, describing a remembered story is different from natural speech, which requires generating or constructing an unconstrained and unscripted message. Description of a remembered story involves recalling and comprehending the fixed content represented in the story, and it requires no process of constructing the internal structures of a message. In fact, Johansson, Holsanova, and Holmqvist (2006) have provided evidence that listening to spatially complex scene descriptions (e.g., a story involving expressions such as at the top, between, to the left of) and retelling them from memory elicits similar eye movement patterns. In other words, retelling the story showed similar cognitive processes as those in understanding the story. Therefore, gestures accompanied by story descriptions may not be the best or most direct evidence to support the idea that language production incorporates active simulations, because gestures could be produced in the recalling and comprehending processes, rather than in producing language.

Although gesture/speech studies have provided valuable evidence for embodied simulation in language production, they are limited to descriptions of remembered
stories. Such production involves simply recalling fixed content from memory (Johansson et al., 2006) and thereby differs markedly from most cases of natural speech production, which typically involve generating the structure of a novel message. Therefore, gestures accompanied by story descriptions may not be the best or most direct evidence to support the idea of active embodied simulations as a critical component of language production.

In short, although gestures can indicate what types of concrete concepts are engaged in performing simulations in concurrent speech, they do not tell us whether or not simulation influences the process of language formulation, including grammatical and lexical encoding processes. More broadly, in addition to the limitation of existing production models discussed in section 3.3, gestures that evidence simulation do not provide a clue to where mental simulation might fit into the system of language production. Thus, the next section addresses the essential question of the existence and designated role of simulations in producing utterances and proposes a new method to empirically investigate the role of motor activities and simulation in sentence production. Experiments in this study are designed to explore the effects of simulation in generating a meaning and transferring it into linguistic materials, instead of in recalling stories. In addition, timing of simulation in relation to message formulation is manipulated in order to carefully examine when simulation interacts with other components in production mechanisms and how. Although the current study aims to investigate simulation effects in sentence production that is as natural as possible, it should be noted that utterances produced with experimental speech constraints in a laboratory setting are not yet fully spontaneous speech. These utterances are different from spontaneous ones especially
because they are single utterances in isolation, with no previous context, and they involve no interlocutors; thus, no communicative purposes are presupposed. In spite of these limitations, the current study aims to move beyond gesture production in recall studies to investigate simulation effects on ongoing message formulation and production.
CHAPTER 4
CLOCK TASK

Experiments 1–8 employed a novel task involving non-linguistic manipulation of bodily motion to investigate such motion’s effects on embodied cognition and simulation, as well as on the message construction process. This chapter describes the task.

4.1 Apparatus

A large mouse pad was created by attaching a black inner circle (size: 20 inches/50 cm in diameter) on top of a white outer circle (size: 24 inches/61 cm in diameter). Both black and white circles were made of thick, solid paper. Four red felt numbers (3, 6, 9, and 12) were glued onto the white circle in the positions in which they are found on a clock (i.e., right, bottom, left, and top of the white circle). Finally, one yellow circle was affixed to the center of the black circle (Figure 4-1). This figure, which resembled a large clock, was placed to the right side of a laptop computer. The relative locations of the equipment are explained in section 4.3.

Figure 4-1: A large clock-patterned mouse pad
4.2 Purpose of using the clock-patterned mouse pad

All experiments designed for this dissertation involved three basic procedures, listed here in random order: participants saw a sequence of images and one of the four numbers (3, 6, 9, or 12), moved a mouse toward that number on the mouse pad, and produced a sentence. This large mouse pad was created to induce smooth hand movement and allow researchers to control participants’ motions without using explicit direction language. For example, showing the number “3” (instead of showing the word right) would direct participants to move their hand to the right because of their clear understanding of the position of a 3 on the right side of a clock face. Since the clock is very familiar in everyday life, participants were able to move their hands in the desired directions without much conscious effort. More importantly, this technique prevented the lexical activation that might result if researchers were to use direction-oriented words such as toward, away, right, and left. This method therefore enabled us to make controlled observations of the effect of physical movements on message formulation. End-of-experiment interviews confirmed that no participant noticed any relationship between the mouse movement and the pictures/speech production.

4.3 Experiment setup

The laptop, microphone, and clock-patterned mouse pad were arranged so as to allow participants to move their hands naturally, and were fixed in place in order to maintain a consistent experimental setting. Eliciting natural hand motion was important in this study because such motion is assumed to evoke schematic experiences that are
associated with toward or away hand motions. Figure 4-2 illustrates the arrangement of the experimental equipment.

![Diagram of apparatus configuration]

(1) Laptop  
(2) Mouse  
(3) Laptop power cable  
(4) Response Box  
(5) Response Box power cable  
(6) Serial cable  
(7) Convert cable  
(8) Microphone  
(9) Clock-patterned mouse pad  
(10) Voice recorder  
(11) Participant

Figure 4-2: Configuration of apparatus
CHAPTER 5

SIMULATION IN UNCONSTRAINED MESSAGE FORMULATION

Great strides have been made in understanding production mechanisms. The models discussed in section 3.2 propose explanations of how speakers transfer the meaning of an intended message into linguistic components. However, it has been difficult to investigate the external factors that influence meaning construction. As described in chapter 2, comprehension research has argued for the importance of embodiment/simulation in sentence interpretation, for example, by showing interactions between motor activities and the comprehension of sentences that describe movement. Moreover, it has been emphasized that the relative timing of simulations and visual perception plays a critical role in either facilitating or hindering comprehension processes.

This chapter presents a set of four experiments. Experiments 1–4 explore the role of embodiment/simulation in sentence production by examining whether or not embodied information activated by direction-specific motor activities is unconsciously incorporated into message formulation processes, specifically when no relationships among the entities in the event are established. The relative timing of action execution and message generation is manipulated to investigate whether variations in timing can affect message formulation.

All four experiments discussed in this chapter involve the “event directionality” of sentences mentioning two objects (e.g., an apple and a basket in the sentence below) that imply movement toward the speaker’s body (“toward sentences,” e.g., I am taking an
apple from the basket) or movement away from the speaker’s body (“away sentences,” e.g., I am putting an apple into the basket). Since object entities are cognitively more basic and isolated and have weaker relational features than actions (Goldin-Meadow, So, Ozyurek, & Mylander, 2008), I predict that motions may easily provide relational information in constructing event representations involving two object entities. In other words, motion may play a significant role in framing event representations and in helping to determine how participants interact with the two object entities.

Section 5.1 presents Experiments 1 (English) and 2 (Japanese), which examine whether prior physical activity affects the determination of event directionality in messages that are not preassociated with any particular relational meaning (for example, in messages involving an apple and a basket, objects that are not preassociated with either toward- or away-movement in speakers’ minds). If nonverbal priming (either toward- or away-motion) activates generalized or previously experienced embodied knowledge, such activation could provide speakers with a basis for the internal structure of the message and dynamically steer the message encoding process, resulting in increased production of the corresponding toward- or away-language.

Section 5.2 describes Experiments 3 (English) and 4 (Japanese), which investigate whether physical motion affects ongoing message construction processes. For example, if seeing pictures of an apple and a basket causes comprehenders to start constructing an internal relationship between them, then are motions still integrated into the ongoing process of message construction as well? This set of experiments may represent real world situations better than Experiments 1 and 2, in which motion execution precedes message formulation, because speakers are likely to engage in physical motion (e.g.,
moving their hands) while generating a message, rather than only moving before thinking.

5.1 Motion before formulating an unconstrained message

Experiments 1 and 2 were identical, with the exception that they focused on, and were conducted in, English and Japanese respectively.

5.1.1 Participants

Participants comprised 40 native speakers of English, all of whom were students at the University of Hawai‘i at Mānoa (UHM), and 47 native speakers of Japanese recruited from UHM, Kapi‘olani Community College, and Tokai University, all in Hawai‘i, as well as Hiroshima University in Japan. They participated in exchange for credit in an introductory linguistics course, a small bag of snacks, or a small amount of monetary compensation. The average time for an entire session was 15 minutes for Japanese and 20 minutes for English. All participants reported normal hearing and vision.

5.1.2 Materials (English, Japanese)

Critical Pictures. The picture stimuli were color drawings adapted from commercially available clip art. Both the English and Japanese experiments utilized identical sets of critical pictures. The critical pictures were grouped into 24 pairs (48 pictures total); one picture in each pair depicted a target object (e.g., a carton of milk), while the other showed a location (e.g., a refrigerator). These pairs were designed so that the relationship between them could be easily described through statements about toward- or
away-motions (e.g., toward-motion in *taking the milk from the refrigerator* and away-motion in *putting the milk into the refrigerator*). Other critical items in the pictures included file and file cabinet, basketball and basketball rack, and plug and outlet (See Appendix A for a complete list of critical pictures). It is important to note, however, that the critical items pictured were not limited to association with each other only through relationships involving toward- or away-motion. For instance, one can *look for the milk in the refrigerator* or *inflate the basketball that is on the basketball rack*. Since no event relationships between the target object and the location were established in advance, the message as well as the syntactic structure was purely unconstrained.

*Filler Pictures.* In the Japanese experiment, an additional 8 pairs of pictures (16 pictures) were included that consisted of a target object (e.g., ring, necklace) and an appropriate body part (e.g., hand, neck). These fillers were created because they were likely to induce toward- (e.g., *putting the ring on a finger*) or away-motion descriptions (e.g., *taking the ring off the finger*), which might boost the use of toward- or away-motion language in critical trials (See Appendix B for a complete list of filler pictures used in the Japanese experiment). They were used as fillers, instead of criticals, because the proportions of toward- versus away-language that they each elicited were strongly biased in favor of one option or the other. Although these fillers worked very well in the Japanese experiment, a pilot study with 17 English speakers showed that they had a strong tendency to use a single verb (specifically, the verb *put*) throughout the experiment. More specifically, the data from ten of the seventeen participants was considered to be strongly lexically primed since they each utilized a single verb in more than 90% of all sentence productions.
throughout the experiment. This lexical priming might occur because the first verbal element in English was always a verb while variables (e.g., DO, PP) were available in the verb final language, Japanese.

In order to avoid lexical priming, a phenomenon that could conceal the effects of motion in constructing event representations, the previously used fillers were discarded and 24 new pairs of filler pictures (48 total filler pictures) were created for the English experiment. The new pairs of objects included pencil and pencil sharpener, wine bottle and wine glass, bat and baseball, and tomato and knife. They were intended to elicit a wide range of verb choices (other than put), such as sharpen, pour, hit, and cut (See Appendix C for a complete list of filler pictures used in the English experiment).

**Picture Norming.** A norming study with eight Japanese and eight English speakers who did not participate in the main experiment verified that each of the 24 pairs of critical pictures could easily induce statements describing toward- or away-motion. Crucially, the proportion of toward-language to away-language in the experimental items was 0.41 to 0.59 in Japanese and 0.44 to 0.56 in English. In other words, a message generated by using two objects could be described in either toward- or away-statements with a similar frequency, regardless of language. This balanced preference for using either toward- or away-expressions maximizes the possibility that the message reflects the effects of an external factor, i.e., motor actions. In addition, the same corresponding groups of participants ensured that the eight pairs of filler pictures used in the Japanese experiment and the 24 pairs of filler pictures utilized in the English experiment easily generated action-related expressions (that were not limited to direction-specific actions).
In this off-line norming test, participants received nine pages of paper with 48 pairs of items for Experiment 1 or five pages of paper with 32 pairs of items for Experiment 2, with the objects presented in a five-column table (See Appendix D for English and Appendix E for Japanese norming tests). The columns included (1) item number, (2) a picture of a target object (e.g., milk, basketball, file), (3) a picture of a location object (e.g., refrigerator, basketball rack, file cabinet), (4) a sentence fragment “I am” in English or “Watashi-wa (I-Nom)” in Japanese, and (5) a figure of a person’s head (        ) whose nose indicated that s/he was looking in a forward direction.

Participants wrote a short description of a possible action involving the two provided objects by completing the sentence fragment “I am” in the fourth column.

The last column was only used for the critical items. For these items, this column included a figure of a person. Participants were asked to draw an arrow (up, down, right, or left) to indicate how the grammatical subject of the sentence (in this case, “I”) would conduct the action about which they had just written. This specification was important because directionality can be ambiguous in sentences. For example, although a prototypical interpretation of the direction implied by the phrase *putting an apple into the basket* is away from the body, the directional interpretation can also be reversed to a toward motion if we assume that the person is putting an apple into a basket that is held close to his body. Note that this concern was unnecessary with almost all items because their locations were fixed (e.g., refrigerator, bookshelf, outlet).

Critical pictures were selected for inclusion in the main experiment if five or more participants in each language group gave responses that fell into the categories of toward- or away-action sentences, as determined by the coding procedures described in section
5.1.5. For filler items, participants were also instructed to create possible action sentences using the two pictures, but were not asked to indicate the direction of the described actions. Filler items were selected if five or more participants provided grammatical responses describing natural actions involving the two objects (e.g., *pouring wine into the glass*, *cutting the kiwi with the knife*, or *eating a cake with the spoon*).

5.1.3 Procedure

The experiment was administered using E-Prime. The participants were tested individually. Participants were seated in front of a computer monitor with the clock-patterned mouse pad (See Chapter 4) placed under the computer mouse. Participants completed a set of 8 practice trials, followed by 48 experimental trials in the English experiment and 32 experimental trials in the Japanese experiment. Each participant was presented with 24 critical trials randomly ordered with 24 (in the English experiment) or 8 (in the Japanese experiment) filler trials.

Participants began with the mouse positioned over the yellow circle that was located at the center of the clock-patterned mouse pad. For each trial, a cross was presented for 500 ms to direct participants’ eyes to the center of the screen, followed by one of four numbers (3, 6, 9, or 12). The number remained on the screen until participants moved the mouse toward the matching number on the clock-patterned mouse pad and left-clicked the mouse upon reaching the number. Participants were asked to keep the mouse on the black surface of the mouse pad; they did not need to actually land on top of the number. They were instructed to hold their hand in that position after clicking, at
which point the sentence fragment “I am…” was displayed for 500 ms followed by two different pictures (displayed one at a time for 300 ms each).

The order of the two pictures was fixed for each language. In the English experiment, a movable target object (e.g., a carton of milk) was followed by a locative object (e.g., refrigerator) (Figure 5-1). This is consistent with English word order, Direct Object-Locative, and was intended to encourage smooth and natural sentence formulation. The order of the two pictures was reversed in the Japanese experiment because Locative-Direct Object is the canonical word order in Japanese (Figure 5-2).

Figure 5-1: Experiment 1 (English). Away motion followed by a sequence of an object picture and a location picture

Figure 5-2: Experiment 2 (Japanese). Toward motion followed by a sequence of a location picture and an object picture

Participants were instructed to complete the sentence starting with “I am” by describing possible actions involving the two pictured objects as quickly as possible, without verbally repeating “I am” and without hesitation (e.g., “uh,” “um”). E-Prime captured the onset of speech, which triggered the replacement of the speaker icon (as
shown in Figures 5.1 and 5.2). This speaker icon screen was intended to notify speakers that their speech was successfully captured by the computer and to motivate them to continue producing their sentence. When participants had finished saying their sentence, they clicked the mouse again, which changed the speaker icon screen to a blank screen. Repositioning the mouse at the center of the mouse pad and clicking would initiate the next sequence. The x-y coordinate positions of the two mouse click events—that is, the one in the center of the pad to initiate the trial and the other near the number on the pad after receiving a number—were recorded into E-Prime. This allowed the researcher to measure and verify the directionality of the hand movement when analyzing the data. Each response sentence produced by participants was separately recorded by a voice recorder (Sony ICD-P520).

The researcher conducted end-of-experiment interviews to discover each participant’s impression of the experiment’s purpose. These confirmed that no participant noticed any relationship between the mouse movement and the pictures/speech production. Most of the participants had no idea about the research’s purpose, and the rest of them incorrectly assumed that measuring the speed of the mouse motions or the accuracy of picture recognition and picture naming were the primary purposes of the experiment.

5.1.4 Conditions

Experiments 1 and 2 had one factor (physical movement) with four levels (directions: toward-, away-, right-, and left-motions), resulting in four conditions. The internal event structure of the two objects was not determined or constrained, but easily
modulated into toward-sentences (e.g., *pulling out the plug from the outlet*) or away-sentences (e.g., *plugging into the outlet*). Therefore, toward- and away-motion conditions, which might generate language denoting corresponding directions, served as critical conditions. On the other hand, left- and right-motion conditions served as the baseline conditions as they would not necessarily facilitate or hinder the generation of sentences associated with toward- or away-directions. These baseline conditions represented the speakers’ underlying preferences for language choice.

Because response types after right-motions and after left-motions showed similar proportions of types of language use (See the response tokens for different language types in section 5.1.7), these two conditions were merged into a single condition (i.e., “left/right condition”) and served as the baseline from which to observe the crucial effects of toward- or away-motions in formulating the subsequent message. As a result, three conditions, toward-, away-, and left/right, were subjected to further statistical analyses.

The experiment was designed as a Latin square, so that each of the 24 pairs of critical pictures appeared in only one motion condition for each participant. Furthermore, each participant received an equal number of picture prompts for each motion condition (i.e., toward-, away-, right-, and left-motion conditions).

5.1.5 Data analysis

Data Coding for Language Type. Totals of 1920 responses from 40 English-speaking participants and 1504 responses from 47 Japanese-speaking participants in the main experiments were recorded. Two undergraduate research assistants (RAs) at the
University of Hawai‘i (one native speaker of English and one bilingual speaker of English and Japanese) and the researcher who is a native speaker of Japanese manually transcribed all the responses that were in their native language into an Excel sheet. From those responses, 960 critical responses in English and 1128 critical responses in Japanese were further analyzed. The research assistants independently examined the linguistic contents of the critical responses and systematically assigned each of them to one of the four categories: toward-language (e.g., taking milk from the refrigerator), away-language (e.g., putting milk in the refrigerator), neutral language (e.g., drinking milk that is stored in the refrigerator), or others (i.e., unanalyzable or ungrammatical sentences). The English-speaking and the bilingual RA made categorical judgments for the English data while the bilingual RA and the researcher did the same for the Japanese data. All coders were blind to the conditions in which responses were produced when they assigned the responses to a particular category. Any disagreements between coders were resolved by discussion between them.

The following criteria were used for all the analyses presented in this dissertation. Sentences involving verbs and/or prepositions that clearly denoted either toward- (e.g., take from) or away-direction (e.g., put into) were labeled as toward- or away-language, respectively. Neutral language involved responses that showed no clear directionality, such as responses describing the state of the event (e.g., looking at an apple in the basket, reading the book that is on the bookshelf, talking to the parrot in the cage), or those that never explicitly stated event directions, but only implied possible motions, such as responses associated with the objects’ primary functions (e.g., heating soup in the microwave, washing a cloth in the washing machine, baking a pie in the oven). More
specifically, for example, *heating soup* might involve a sequence of actions: holding a bowl of soup, opening and putting the soup in the microwave, closing the microwave, and setting the timer. However, in a case like this, the motions were only implied and were too complicated to be analyzed.

Finally, responses were categorized as “others” (1) when trials were skipped or incomplete/ungrammatical sentences were produced, (2) when sentences did not include both of the two pictured objects, or (3) when articulated sentences denoted clear directions, but they involved multiple actions with opposite directions. For example, the phrase *taking a shirt and putting it onto my clothesline* implies a toward-motion (i.e., *taking a shirt*) followed by an away-motion (i.e., *putting it onto my clothesline*). However, if a response described multiple actions representing a single direction, it was analyzed based on that direction (e.g., toward motion implied by a sentence *I ate an apple after selecting the best one from the fruit basket*).

**Data Coding for Word Order in Japanese.** Although the order of two noun phrases, namely, the Locative and the DO, can be flexibly positioned in Japanese due to various linguistic or discourse factors (See section 3.5), the canonical word order is (Subject)-Locative-Direct Object-Verb (Ferreira, 1996; Ferreira & Yoshita, 2003; Nakano, Felser, & Clahsen, 2002, Yamashita & Chang, 2001). The locative phrase, which is marked with a locative marker (e.g., -kara ‘from’, -ni ‘into’), denotes a specific directionality (e.g., toward- or away-motion, respectively) while the direct object, marked with the accusative marker –o, denotes no specific direction. For each response in the Japanese experiment, two coders (one of the RAs and the researcher) examined the word order and labeled
them as one of the three types, namely, canonical (Locative phrase first, Direct Object second), scrambling (Direct Object first, Locative phrase second), or neither word order (e.g., a relative clause such as [kago-nonaka-ni oite-atta] ringo-o tabeta; gloss: [basket-inside-Loc place-past] apple-Acc ate; ‘ate an apple that was placed in the basket’).

**Accuracy for Mouse Movement.** E-Prime was programmed to record the x-y coordinate positions of the two mouse click events—that is, the one in the center of the pad to initiate the trial and the other near the number on the pad after receiving a number. All experiments (Experiments 1–8) employed these x-y coordinate positions of the two mouse click events to examine the accuracy of the directionality of the hand motion.

**Selected Participants.** Several criteria were employed to decide which data were valuable for further analysis. First, the accuracy of participants’ mouse movements was calculated based on the entire set of trials (including critical and filler items), with the idea that participants should be excluded if their accuracy rate was below 70%. One English-speaking participant with mouse inaccuracy of 69% was removed (the average accuracy from 40 people was 88%). No participants were removed due to mouse inaccuracy from the Japanese experiment (the average accuracy from 47 people was 88%). Second, responses with incorrect mouse movements for each participant were eliminated. This resulted in less than 11% of the English data and less than 12% of the Japanese data being removed. Third, participants exhibiting lexical persistence were excluded. Lexical persistence was defined as a participant’s strong tendency to use a single verb throughout the experiment although it might produce unnatural or ungrammatical utterances. No
participant from the English experiment was excluded, and one participant from the Japanese experiment was excluded due to lexical persistence. Finally, participants who did not understand the experimental task (e.g., those who produced utterances including the sentence fragment “I am” or mouse clicks at incorrect times) were also eliminated. No English and two Japanese speakers were omitted for this reason. As a result, 39 English speakers and 44 Japanese speakers were left. In order to have an equal number of participants for each condition, 36 English speakers (i.e., nine participants for each of four conditions) and 36 Japanese speakers (i.e., nine participants for each of four conditions) were selected for further analysis.

5.1.6 Predictions

Predictions for Language Content. Motions executed immediately before formulating the message are expected to activate experiential knowledge associated with those particular motions. Since the internal relations between the two pictured objects were not determined (but the relational meaning is one of the crucial components in framing a message as described by the production model in section 3.2), previously activated schematic knowledge may be rapidly used as an event structure frame. Direction-oriented activities feed directional information to the subsequent process of meaning construction, which will be used to configure the relation of the objects and the event structures in the message. If a message that is associated with no relational meaning is susceptible to body actions, engaging in toward-movement will increase the proportion of sentences denoting toward-movement (e.g., taking an apple from the basket) while away-movement will facilitate the production of away-language (e.g., putting an apple into the basket).
Because the message at the message planning stage is not linked with any linguistic materials, no significant differences should be observed in two typologically different languages, that is, English (Experiment 1) and Japanese (Experiment 2). Therefore, previous motions in a particular direction are predicted to foster speech output denoting that same direction both in English and Japanese.

**Predictions for Word Order in Japanese.** The flexible word order of the Japanese messages can be an ideal mediator to reflect motor effects on message formulation (See section 3.5 for detailed discussion). This dependent measure is especially important for providing transparent evidence for the isomorphic relations between phrasal order and simulated event representations. More specifically, simulating toward-events drives a speaker’s mental focus to identify the location (Loc) first followed by the target object (DO), resulting in an increased number of Loc-DO sequences in toward-language. In contrast, simulating away-events causes a speaker to mentally focus on the target object (DO) and then spot the target location (Loc); therefore, a word order consistent with the temporal sequence of the simulated event, i.e., DO-Loc, is preferred in describing away-events. In short, the word order can indicate a “mental trace” of the speaker’s message constructions.

5.1.7 Results

Table 5-1 summarizes the tokens of critical responses for three conditions in each language. Note that the condition labeled “Left/Right Motion” (which served as the baseline language preference) is the average of two neutral/baseline conditions, namely,
left-motion and right-motion conditions. The averages are presented in the table while the numbers in parentheses are the combined tokens from those two conditions.

Table 5-1: Number of productions with Away, Toward, Neutral, or Other responses in the motion-first, picture-second design:

<table>
<thead>
<tr>
<th>Sentence Type</th>
<th>Experiment 1 (English)</th>
<th>Experiment 2 (Japanese)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Away Motion</td>
<td>Toward Motion</td>
</tr>
<tr>
<td>Away Lg. (Canon.)</td>
<td>145 [71.1%]**</td>
<td>119 [69.2%]</td>
</tr>
<tr>
<td>Away Lg. (Scram.)</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Toward Lg. (Canon.)</td>
<td>13 [6.4%]</td>
<td>10 [5.8%]</td>
</tr>
<tr>
<td>Toward Lg. (Scram.)</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Neutral Lg.</td>
<td>39 [19.1%]</td>
<td>29 [16.9%]</td>
</tr>
<tr>
<td>Other Lg.</td>
<td>7 [3.4%]</td>
<td>14 [8.1%]</td>
</tr>
<tr>
<td>Total</td>
<td>204 [100%]</td>
<td>172 [100%]</td>
</tr>
</tbody>
</table>

*The number in the ( ) indicates the tokens from both left-motion and right-motion.

**Percentages out of trials in the designated condition are indicated in the [ ].

5.1.7.1 English speakers (Experiment 1)

Since participants did not repeat the given sentence fragment “I am…” the first English verbal component after the physical motion was a verb, which encoded event directionality (e.g., toward-motion indicated by pull out or take out; away-motion implied by plug in or put in). This order of motor execution and verb selection crucially allowed us to clearly examine whether or not direction-specific motions could influence the event encoding process, as no other element intervened between motion and verb selection.

One-way repeated measures ANOVA with number of away responses as the dependent measure and three levels of the prime motion type (toward-, away-, left/right-motions) as independent measures showed significant effects of motion on language
choice ($F_1 (2, 70) = 4.2, p < .05; F_2 (2, 46) = 5.1, p < .05$) (See Figure 5-3 for the proportions of response types for each motion condition). Paired t-tests conducted on away responses revealed that the number of away responses was increased significantly after conducting away-motion than after conducting toward-motion ($t_1 = 2.3, p < .05; t_2 = 3.1, p < .01$), but the difference in away responses between away-motion and left/right-motion conditions was not significant ($t_1 = .74, p = .47; t_2 = .62, p = .54$). In contrast to my prediction, one-way repeated measures ANOVA with number of toward responses as the dependent measure showed no significant effects of motion on language choice in either participant or item analyses ($F_1 (2, 70) = 2.6, p = .083; F_2 (2, 46) = .6, p = .56$). However, this analysis is limited by the small number of toward-language responses, and this may explain why no significant effect is observed in toward-language responses. These results generally suggest that event language chosen only for away-message, but not for toward-message, was influenced by hand motions executed prior to the message construction in English. The data consistently shows an overall away-language bias, and plausible causes of this asymmetry between away- and toward-language frequency will be discussed in section 5.1.8.
The following Table 5-2 provides typical responses for away, toward, neutral, and other language.

Table 5-2: Sample utterances for each type of language category in Experiment 1

<table>
<thead>
<tr>
<th>Responses</th>
<th>Away language</th>
<th>Toward language</th>
<th>Neutral language</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Putting the apple in the fruit basket.</td>
<td>Taking my glasses out of their case.</td>
<td>Washing clothes in the washing machine.</td>
<td>Microwaving the soup.</td>
</tr>
<tr>
<td></td>
<td>Plugging the cord into the outlet.</td>
<td>Grabbing the pen from the pen holder.</td>
<td>Looking at a parrot in the cage.</td>
<td>Decorating the Christmas tree.</td>
</tr>
<tr>
<td></td>
<td>Hanging the hat on the rack.</td>
<td>Getting a milk from the refrigerator.</td>
<td>Baking a pie in the oven.</td>
<td>Reading the book and then putting it on the bookshelf.</td>
</tr>
</tbody>
</table>

Figure 5-3: Percentage of English responses after performing motion followed by picture perception
5.1.7.2 Japanese speakers (Experiment 2)

As with the English data, toward and away responses in Japanese were separately analyzed by one-way repeated measures ANOVAs in terms of subjects and items. As predicted, toward responses clearly showed significant effects in both participant and item analyses ($F_1(2, 70) = 29$, $p< .001$; $F_2(2, 46) = 46.6$, $p<.001$). Further, paired t-tests conducted on toward responses revealed that toward responses after toward-motion were significantly greater than those after unrelated (left/right) motion ($t_1=6.4$, $p<.001$; $t_2=7.5$, $p<.001$) or away-motion ($t_1=6.6$, $p<.001$; $t_2=8.9$, $p<.001$).

Likewise, the one-way repeated measures ANOVA was conducted on away responses and showed significant effects of motion on direction-related language choice in subject and item analyses ($F_1(2, 70) = 53.2$, $p< .001$; $F_2(2, 46) = 76.6$, $p<.001$). Paired t-tests indicated that speakers were more likely to produce away-language after away-motion than after left/right-motion ($t_1=4.3$, $p<.001$; $t_2=4.2$, $p<.001$) as well as after toward-motion ($t_1=8.1$, $p<.001$; $t_2=11.5$, $p<.001$). A significant difference was consistently observed between the number of toward or away responses after unrelated (left/right) motion and those after incompatible motions (away-motion before a toward response or vice versa).
Table 5-3 illustrates typical toward, away, neutral, and other responses with canonical and scrambled word orders.

<table>
<thead>
<tr>
<th>Language Type</th>
<th>Word Order</th>
<th>Responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Away language</td>
<td>Canonical</td>
<td>Yoohukukake-ni yohuku-o kakeru cloth rack-Loc cloth-Acc hang hang the cloth on the cloth rack.”</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pentate-ni pen-o modosu penholder-Loc pen-Acc return Return the pen into the penholder.”</td>
</tr>
<tr>
<td></td>
<td>Scrambled</td>
<td>Suupu-o denshirenji-ni ireru soup-Acc microwave-Loc put Put the soup in the microwave.”</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Huraipan-o konro-ni noseru. pan-Acc portable cooking stove-Loc put Put the pan on the portable cooking stove.”</td>
</tr>
</tbody>
</table>
Word order is examined in order to illuminate two related questions: (a) do speakers tend to select a particular word order that matches the temporal sequence of the simulated event? and (b) among those responses that show event–word order consistency, what proportions of word order selection are directly influenced by motor activities?

The first analysis attempts to answer the question of whether word order reflects the temporal order of simulated events. If it does, utterances denoting away-motion should be more likely to be expressed in scrambled word order (because an agent pays attention to an object and then identifies the location, that is, the object’s destination). On the other hand, utterances denoting toward-motion should be expressed more often in scrambled word order (because an agent mentally confirms the location and then

<table>
<thead>
<tr>
<th>Toward language</th>
<th>Canonical</th>
<th>Scrambled</th>
<th>Neutral language</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wainnotana-kara wain-o toru wine rack-Loc wine-Acc take “Take a wine bottle from the wine rack.”</td>
<td>Hoosekibako-kara yubiwa-o dashita jewelry box-Loc ring-Acc let out “Take out the ring from the jewelry box”</td>
<td>Tori-o torikago-kara toridasu bird-Acc cage-Loc take out “Take out the bird from the cage.”</td>
<td>Sentakuki-de tao-o arau washing machine-Loc towel-Acc wash “Wash the towel in the washing machine.”</td>
<td>CD-o kiku CD-Acc listen “Listen to a CD”</td>
</tr>
<tr>
<td>Kagi-o kagiire-kara toru key-Acc key box-Loc take “Take the key from the key box.”</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
approaches the object). This relationship between word order and the sequence of simulated events was examined only in toward- and away-language since the event directionality of neutral and other language was unclear or too complex to select a single direction. For each of the toward- and away-language responses, proportions of canonical and scrambled word order were calculated for each participant and for each item. For example, in toward-language, proportions of canonical word order (calculated by counts of canonical responses in toward-language/sum of canonical responses in toward- and away-language) and proportions of scrambled word order (calculated by counts of scrambled responses in toward-language/sum of scrambled responses in toward- and away-language) were compared by using paired t-tests. These paired t-tests assessed whether the word order was influenced by the choice of toward-language vs. away-language.

As predicted in 5.1.6, paired t-tests revealed a significant relationship between event language and word order: away-language sentences were more likely to have scrambled word order than canonical word order ($t_1=3.2$, $p<.005$; $t_2=5.6$, $p<.001$), while toward-language sentences were expressed more in canonical than in scrambled word order ($t_1=4.9$, $p<.001$; $t_2=5.6$, $p<.001$). The strong relationships observed between the event type (i.e., toward- vs. away-language) and word order (i.e., canonical vs. scrambled word order, respectively) indicate that speakers mentally simulate and experience the event in chronological order and that mental activities affect word order.

The second question is whether physical motion affects the word order. To answer this question, it should be noted that there is an intermediate stage between motion and word order selection, that is, an event determination. If production processes
are sensitive to the speakers’ physical status, then motor actions should influence the message/event structure, and this ultimately affects the word order. If motion is the causal component that eventually affects word order, then the proportions of utterances using canonical word order to express the toward-event should be the highest after toward-motion enactment (compared to those found after left/right- or away-motion). In the same line of logic, the probabilities of utilizing scrambled word order for utterances involving away-motion should be the higher after compatible (away) motion than after unrelated (left/right) or opposite (toward) motion enactments.

As Figure 5-5 illustrates, with toward-language, speakers employ a larger number of canonical word order sentences after toward-motions (i.e., 44.3% or 58 tokens) compared to after left/right- (i.e., 22.1% or 33 tokens) or away-motions (i.e., 14.8% or 25 tokens). As expected, with away-language, the number of scrambled word order sentences is significantly higher after away-motion (i.e., 49.1% or 83 tokens) than after left/right- (i.e., 39.6% or 59 tokens) or toward-motion (i.e., 17.6% or 23 tokens).
These word order results suggest that word order is affected by the chronological order of simulated events. Moreover, such consistency between event type and word order is maximized when the compatible motor activities are conducted. In other words, speakers selectively utilize a particular word order by accommodating not only event types but also motor information.

5.1.8 Discussion

The above results generally suggest that motor actions evoke action-related or direction-specific information and leave experiential traces in the cognitive faculty. When formulating a message by organizing or arranging two objects’ event representations, the language processors unconsciously accept the recently activated directional information and utilize it in framing the subsequent message. This reveals that the language
production mechanisms are sensitive to non-linguistic, physically grounded information at the message-generating stage, and such physically based information dynamically modulates the subsequent message when no specific event structure is associated with the message.

There is appreciable asymmetry between away- and toward-language frequency in the English data throughout the conditions (e.g., away-language after away-, toward-, and left/right-motion: 71.1%, 69.2%, and 71.3%; toward-language after away-, toward-, and left/right-motion: 6.4%, 5.8%, and 6.9%) compared to the Japanese data, which shows more balanced frequency between toward- and away-language (e.g., away-language after away-, toward-, and left/right-motion: 63.9%, 30.3%, and 54.7%; toward-language after away-, toward-, and left/right-motion: 16.1%, 49.1%, and 22.5%). The English data consistently shows a strong bias for away-language in which speakers may mentally interact with objects in a particular order, that is, attention focuses on the direct object (e.g., *an apple*) first and then on a location (e.g., *a refrigerator*) (O’Grady & Lee, 2005). If speakers are engaged in creating internal event structures in mental simulation, the order of the picture presentations—for example, when a movable object (e.g., *milk*) is subsequently followed by a location (e.g., *a refrigerator*)—may influence the message/event generation processes. In other word, speakers prefer to produce utterances involving away-motion because speakers first perceive and activate a concept of direct object and then a concept of location, in an activation whose temporal order is coherent with the temporal order of the event in the real world.

Furthermore, that no motion effects on toward-language were observed in English might also be due to this stronger underlying preference for away-language (percentages
of away-language vs. toward-language after left- or right-motion: 91.1% vs. 8.9%) than that observed in Japanese (percentages of away-language vs. toward-language after left- or right-motion: 70.8% vs. 29.2%). This language bias underlying English might be too strong to be overridden or influenced by other external factors, such as motion. In addition, different temporal points in English and Japanese to determine the event structure might cause non-significant effects in English. In the head final language Japanese, which allows a relatively flexible word order, the initial verbal argument in this experiment could be either a locative phrase (in canonical word order) or a direct object (in scrambled word order). In the case of responses with a non-canonical, scrambled word order, although the accusative case marker (-o) attached to the direct objects does not aid in anticipating the upcoming direction-specific event, variable locative case markers (e.g., -kara or -ni) imply the general directionality (e.g., toward-, or away-motions, respectively) of the motor action represented in the event. Unlike English, the event types/directionalities are not finalized until the locative phrase is produced, and this extra time or flexibility in creating event structures may expand the possibility of incorporating the motor information into the process of relational meaning constructions.

The word order that speakers select to describe an event with a specific direction indicates the manner in which they are simulating that event while simultaneously producing language. Speakers prefer to produce linguistic orders that align with the spatial and temporal order of events in the real world (Goldin-Meadow et al., 2008; O’Grady & Lee, 2005): in away-sentences, attention centers first on the direct object and then on a location that would typically be further away from the agent (resulting in more DO-Loc sentences), while in toward-sentences, the typically distal locations are
mentioned before the direct objects (resulting in more Loc-DO sentences). This consistency of Japanese word order with temporal reality supports the claim that speakers mentally access event representations while producing language.

5.2 Motion while formulating an unconstrained message

This section introduces one experiment in English (Experiment 3) and one in Japanese (Experiment 4). These are identical to Experiments 1 (English) and 2 (Japanese), except that the physical activity is performed after perceiving the sequence of pictures. Participants first see the “I am…” screen followed by the sequence of two pictures. After the pictures, they see one of four numbers (3, 6, 9, or 12) on the screen. Just as in the previous experiments, participants move the mouse in the direction indicated by the number displayed on the screen. While Experiments 1 and 2 examined motor actions’ influence on the subsequent process of meaning construction, this set of experiments investigates whether or not motor actions can be integrated into the ongoing process of message construction.

This modification of the perception-motion timing allows us to explicate the role of bodily movements in the process of language formulation. If language processors continuously receive action as influential external information and attempt to utilize it in configuring internal objects’ relations even though the direction-oriented event structure has emerged already, then results similar to previous experiments (i.e., motor actions’ significant influence on meaning construction) should be observed in these experiments. On the other hand, if language processors can utilize the directional information provided by physical activities only when the objects’ relation is not yet determined or associated
with a particular event structure, then no motor influence on meaning construction should be observed in Experiments 3 and 4 because the determination of internal relations may have already started at the time of the physical movement.

5.2.1 Participants (English, Japanese)

Forty-eight native speakers of English who were students at the University of Hawai‘i and 40 native speakers of Japanese who were students at the University of Hawai‘i, Kapi‘olani Community College, or Hiroshima University participated in exchange for credit in introductory linguistics courses, a small bag of snacks, or a small amount of monetary compensation. All participants reported normal hearing and vision.

5.2.2 Materials (English, Japanese)

The materials used in Experiments 3 (English) and 4 (Japanese) were identical to those used in Experiments 1 and 2, respectively.

5.2.3 Procedure

The procedures in Experiments 3 (Figure 5-6) and 4 (Figure 5-7) were identical to those in Experiments 1 and 2, except that the object-location pictures were displayed prior to the directional prompt. Therefore, first, participants were shown a screen with the sentence fragment “I am…” for 500 ms followed by two pictures in rapid succession (300 ms for each). A cross then appeared on the center of the screen for 500 ms and then was replaced by one of the four numbers 3, 6, 9, or 12. As soon as participants moved the mouse near the matching number on the mouse pad and clicked, the screen went blank.
Participants completed a sentence starting with “I am” that described a possible action involving the two pictured objects, without verbally repeating “I am.” The onset of their verbal responses triggered the appearance on the screen of a speaker icon, which was intended to encourage participants’ speech continuation. Following sentence completion, participants clicked the mouse, repositioned it back to the center of the mouse pad, and clicked again. Participants then repeated these procedures for the next trial.

5.2.4 Conditions

Just as in Experiments 1 and 2, Experiments 3 and 4 have four conditions, comprised of one factor (physical movement) with four levels (the directions: toward-, away-, right-, and left-motions).
5.2.5 Data Analysis

Experiments 3 and 4 followed the same criteria for coding the language type and word order as those used in Experiments 1 and 2.

**Data Coding for Language Type.** Totals of 2304 responses from 48 English-speaking participants and 1280 responses from 40 Japanese-speaking participants in the main experiments were recorded. Further analyses were conducted on only 1152 critical English responses and 960 critical Japanese responses.

**Accuracy for Mouse Movement.** As stated in Section 5.1.5, the x-y coordinate positions of the two mouse click events were recorded and used to examine the mouse movement accuracy.

**Selected Participants.** No English participant and three Japanese participants were excluded because of not understanding the experimental task. No indication of lexical persistence was observed in either language. One English-speaking participant, whose average mouse accuracy across all trials was 58%, was removed due to low (under 70%) mouse accuracy. In order to have an equal number of participants for each list, three participants were eliminated, resulting in 44 participants in total in the English experiment. The average accuracy for 44 English participants was 88.4%. No participants were eliminated from the Japanese experiment due to low mouse accuracy; their average accuracy was above 70% for all trials (i.e., 87.1% for 40 native Japanese speakers).
Responses with incorrect mouse movements were eliminated, which resulted in the removal of less than 11.6% of the entire data in English (i.e., 245 incorrect movements out of 2112 movements) and less than 13.1% in Japanese (i.e., 168 incorrect movements out of 1280). As a result, 44 native English speakers (i.e., 11 participants for each of four conditions) and 40 native Japanese speakers (i.e., 10 participants for each of four conditions) were selected for further analyses.

5.2.6 Predictions

Predictions for Language Content. Since participants were informed ahead of time that their task was to create a sentence by using two pictured objects, the message planning process should have started immediately after they perceived the pictures. When participants executed the physical movement, they should have already completed or been in the process of completing their message planning. If the message planning process was sensitive to concurrent motions and able to utilize the directional information given by these motions in configuring event representations, then such directional information would be immediately integrated into the linguistic processes and reflected in the linguistic content. That is, toward-motion would facilitate the use of sentences denoting toward-motion while away-motion would increase the production of away-motion sentences. Since direction-related experiential knowledge activated by physical motion was introduced into the message planning period, similar motor effects on message modulation should be observed in English and Japanese, regardless of linguistic or structural differences.
**Predictions for Word Order in Japanese.** Based on the same reasons discussed in section 5.1.6, I predicted that event descriptions involving toward-directional actions would be more likely to be produced in canonical word order (Loc-DO), while those denoting away-directions would be more likely to be represented in scrambled word order (DO-Loc).

### 5.2.7 Results

Table 5-4 summarizes the raw counts of critical responses for three conditions in English and Japanese.

<table>
<thead>
<tr>
<th>Sentence Type</th>
<th>Experiment 3 (English)</th>
<th>Experiment 4 (Japanese)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Away Motion</td>
<td>Toward Motion</td>
</tr>
<tr>
<td>Away Lg. (Canon.)</td>
<td>153 [61.7%]**</td>
<td>76 [34.5%]</td>
</tr>
<tr>
<td></td>
<td>73 [33%]</td>
<td>36 [21.8%]</td>
</tr>
<tr>
<td>Away Lg. (Scram.)</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>70 [31.7%]</td>
<td>26 [15.8%]</td>
</tr>
<tr>
<td>Toward Lg. (Canon.)</td>
<td>23 [9.3%]</td>
<td>66 [30%]</td>
</tr>
<tr>
<td></td>
<td>39 [17.6%]</td>
<td>47 [28.5%]</td>
</tr>
<tr>
<td>Toward Lg. (Scram.)</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>8 [3.6%]</td>
<td>28 [17%]</td>
</tr>
<tr>
<td>Neutral Lg.</td>
<td>63 [25.4%]</td>
<td>69 [31.4%]</td>
</tr>
<tr>
<td></td>
<td>27 [12.2%]</td>
<td>22 [13.3%]</td>
</tr>
<tr>
<td>Other Lg.</td>
<td>9 [3.6%]</td>
<td>9 [4.1%]</td>
</tr>
<tr>
<td></td>
<td>4 [1.8%]</td>
<td>6 [3.6%]</td>
</tr>
<tr>
<td>Total</td>
<td>248 [100%]</td>
<td>220 [100%]</td>
</tr>
<tr>
<td></td>
<td>221 [100%]</td>
<td>165 [100%]</td>
</tr>
</tbody>
</table>

*The number in the ( ) indicates the tokens from both left-motion and right-motion.

**Percentages out of trials in the designated condition are indicated in the [ ].
5.2.7.1 English speakers (Experiment 3)

One-way repeated measures ANOVAs with number of toward and away responses as the dependent measures with three levels of prime motions (e.g., toward-, away-, or left/right-motions) showed significant effects of motion on language both in the subject and item analyses (i.e., toward-language: $F_1(2, 86)=20.6, p< .001; F_2(2, 46) = 47.8, p< .001$ and away-language: $F_1(2, 86)=34.4, p< .001; F_2(2, 46) = 30.2, p< .001$) (See Figure 5-8). Paired t-tests showed that both toward and away responses were significantly increased by the corresponding motion prime compared to neutral (left/right) motions (in toward responses $t_1=5.3, p<.001; t_2= 7.9, p<.001$; in away responses $t_1=2.5, p<.05; t_2= 2.3, p<.05$) and compared to incompatible motions (in toward responses $t_1=4.7, p<.001; t_2= 8.6, p<.001$; in away responses $t_1=7.0, p<.001; t_2= 6.2, p<.001$). These results showed that toward- and away-motor actions made a significant impact on the concurrent message formulation process and influenced the subsequent linguistic encoding processes.
Figure 5-8: Percentage of English responses after perceiving pictures followed by motion

5.2.7.2 Japanese speakers (Experiment 4)

As predicted, one-way repeated measures ANOVAs with number of toward and away utterances as the dependent measures showed significant effects both in participant and item analyses (i.e., toward-language: \(F_1(2, 78) = 5.3, p < .01\), \(F_2(2, 46) = 8.0, p < .01\), and away-language: \(F_1(2, 78) = 40.6, p < .001\); \(F_2(2, 46) = 45.7, p < .001\)) (See Figure 5-8). Paired t-tests revealed that both toward and away responses were significantly increased by the associated toward- and away-motion primes compared to neutral (left/right) motions (in toward responses \(t_1 = 2.8, p < .01\); \(t_2 = 2.8, p < .05\); in away responses \(t_1 = 4.7, p < .001\); \(t_2 = 4.4, p < .001\)) and compared to incompatible away- or toward-motions, respectively (in toward responses \(t_1 = 2.4, p < .05\); \(t_2 = 3.3, p < .005\); in away responses \(t_1 = 8.5, p < .001\); \(t_2 = 9.1, p < .001\)). The results suggested that motor activities dynamically influenced the ongoing determination of event structure, that is, toward- or away-movements are able to influence message formation even after event planning has started.
As described in section 5.1.7.2, two separate analyses were conducted to investigate the relationships among motion, event types, and word order. First, paired t-tests were conducted to statistically examine whether speakers manipulate word order when expressing toward- or away-events constituted by a particular temporal sequence. As predicted, speakers were more likely to use canonical word order than non-canonical word order when expressing toward-events ($t_1=4.0$, $p<.001$; $t_2=3.4$, $p<.005$), although the word order effects were significant in producing away-events by item analysis ($t_2=3.4$, $p<.01$), but diminished by subject analysis ($t_1=1.0$, $p=.3$).

The second question is what portion of the word order selection, among the responses that show a consistency between the temporal order of the event and word order, is directly affected by physical motion? If motion in fact influences the message construction as well as word order determination, then canonical word order should be
produced the most when expressing events involving toward-motion (i.e., toward-language) after toward-motion enactment. Likewise, scrambled word order should be observed the most in describing away-events after away-motion execution. As Figure 5-10 shows, speakers utilize canonical word order the most in expressing toward-language when toward-motion is engaged (i.e., toward-language after toward-, left/right-, and away-motion: 34.3% or 47 tokens, 25.2% or 40 tokens, and 20.5% or 39 tokens). Moreover, speakers tend to use scrambled word order the most in expressing away-events when away-motion is performed (i.e., away-language after toward-, left/right-, and away-motion: 19.0% or 26 tokens, 33.4% or 53 tokens, and 36.8% or 70 tokens).

Figure 5-10: Counts and percentages of toward and away responses with canonical or scrambled word order after performing motion in specific directions (Experiment 4)

In short, not only simulations (evoked during message/event formulation), but also motor actions (performed in response to the number on the screen) significantly influence the selection among the flexible word orders in Japanese.
5.2.8 Discussion

Results for each language showed that motion also affected the ongoing message construction processes, resulting in influencing toward or away predicate choice, although the word order effect was partially diminished. The incomplete (or weak) word order effect might indicate that participants were already formulating a message using two objects when direction-specific information (driven by hand motion) intervened in the meaning construction process. As direction information from the external influence became more prominent in the speakers’ cognitive status, the language processors attempted to incorporate this new information into ongoing meaning construction. As a result, the physical motion successfully influenced verb selection in sentence-final position, but it came too late to be fully reflected in word order.

The observation of similar results for language choices in Experiments 3 and 4, despite the differences between the two languages used, strengthens the idea that ongoing message construction systematically incorporated directional information activated by physical motion and that this was reflected in the upcoming speech. Things that distinguished this set of experiments from Experiments 1 and 2 were that the motor effects on message construction associated with toward- and away-motions were consistently observed in both English and Japanese experiments. This might indicate that our production mechanisms are more prepared to integrate motor information after message formulation has taken place because the language processors that are actively generating a message attempt to decide the relational information.
5.3 General discussion

The experiments presented in this chapter examined whether or not motions grounded in our experiences can influence subsequent (Experiments 1 and 2) or ongoing (Experiments 3 and 4) message planning. The results from language choices (i.e., toward versus away content) indicate that direction-oriented activities moderately feed directional information to subsequent (Experiments 1 and 2) and effectively manipulate concurrent (Experiments 3 and 4) processes of meaning construction, which in turn are used to configure the relative roles of the objects involved in the events being described.

Findings of significant motion effects on utterances in Experiments 1 and 2 showed that body movements (e.g., toward- or away-motions) activated accumulated/schematic knowledge from previous cognitive experiences. Speakers used the experiential domains activated by non-linguistic priming as an aid in planning the internal structure of unconstrained messages, if they were not biased or fixed to use a particular type of event inflexibly. This implied that the message or thought was derived not only from the speaker’s intention or ideas, but also from non-linguistic, physical motions that were previously performed. Unconstrained messages were receptive to and influenced by directional information that was stimulated by external motion.

Observed motion effects on utterances in Experiments 3 and 4 indicated that speakers were concurrently capturing the immediate motions they were engaged in and incorporating them into the language generation process in the course of message formulation. Speakers were indeed more sensitive to their physical status when motion was executed while constructing a message than before constructing a message. These results suggest two plausible assumptions regarding the role of physical motion in
sentence production. First, if messages are already assigned to particular linguistic entities at the moment of motion enactments, then the results indicate that physical motion forcefully affects and overrides the linguistic choices. Because this set of experiments is designed to have visual representations for 1100 ms (i.e., two picture representations for 300 ms each followed by a cross for 500 ms) before speakers perform motor activities, this may be long enough for them to start generating a message and associating it with particular linguistic materials. Second, if messages are not yet associated with linguistic entities when motion is conducted, the results in Experiments 3 and 4 indicate that motor information can be immediately incorporated into the message construction.

In order to elucidate the linguistic status of the message when motor components come into play, the sizes of the motor effects on linguistic choices in Experiments 1 and 3 are compared. If the first assumption is true, and messages that are already linguistically encoded can still be modulated by integrating motor information, then in Experiment 3 the motor effects on language should be smaller than in Experiment 1, where initially activated motions are integrated into subsequent messages that are not associated with any particular linguistic materials. However, significant motion effects are only observed in Experiment 3, whereas they are relatively weak and inconsistent in Experiment 1. This comparison supports the second assumption, indicating that when messages or thoughts are not yet associated with any linguistic materials when speakers perform body motions, then language processors promptly integrate motor information immediately before utterances.
The combination of Experiments 1–2 and 3–4 implied that the language processors are vulnerable to the speakers’ previous and concurrent body motions, not only at the initial message planning stage, but also during ongoing meaning creation when a message is unconstrained so that its relational meaning is not established.

This susceptibility of production mechanisms, that is, that the language processors are sensitive to previously activated information, is also observed in the well-investigated phenomenon known as *syntactic priming*. Syntactic priming is the process by which the syntactic structures that speakers utilize implicitly affect their subsequent speech (Bock, 1986; Bock & Loebell, 1990). Speakers have been shown to prefer a syntactic structure that they have recently used over alternate constructions. For example, double object (e.g., *Mary gave me a letter*) and prepositional dative structures (e.g., *Mary gave a letter to me*) are semantically similar but syntactically different constructions. Similarly, active and passive constructions can both be used to describe a single event (e.g., *Lightening is striking a church* vs. *The church is being struck by lightening*; Bock & Warren, 1985). In cases such as these, when speakers have more than one syntactic structure available to them, they are more likely to choose, unconsciously, the same structure as they used in previous utterances. Susceptibility to syntactic priming is one of the fundamental characteristics of sentence production mechanisms. This phenomenon is not due to a particular task dependence or a specific task demand because syntactic persistence is successfully observed even when a wide range of methodologies are used, including picture description (Bock, 1986; Bock & Warren, 1985), sentence recall (Ferreira & Yoshita, 2003), and sentence completion in written and spoken language (Pickering & Branigan, 1998; Yamashita & Chang, 2001).
The word order preferences in Japanese (Experiments 2 and 4) that reflect the chronological order of the events represented support the claim that speakers were engaged in mental simulation and were re-experiencing the event while producing language. Crucially, such compatibility between event type and word order is dramatically observed specifically when motion is enacted in a consistent direction. This supports the idea that motion plays an influential role in determining the event structure that eventually influences the word order. Since the motions used here were typical motions that could easily co-occur with everyday speech, this provides groundwork for future investigations of how motions might influence message generation in fully spontaneous speech.

In short, this chapter showed significant motor effects on message formulation when no linguistic material was encoded. More specifically, messages in which the objects’ roles in the event structure were not fully established (Experiments 1 and 2) and in which the objects’ roles in the event structure might be in the process of establishment, but not yet be associated with linguistic material (Experiments 3 and 4) could voluntarily integrate previously activated or concurrently salient directional information into meaning construction. These sets of results that show motor effects on (a) messages with no relational meaning associated and (b) messages whose relational meaning may be being formulated raise the following question: does motion influence fixed messages whose relational meaning is completely established?

Therefore, the next chapter will expand the tested linguistic types to include constrained/fixed messages (i.e., messages with an established internal structure). The inclusion of fixed messages addresses the fundamental question of whether or not motor
action can play an influential role in producing language about fixed messages (Experiments 5–8) in a similar manner as that observed in producing language about unconstrained messages (Experiments 1–4). If motion is an optional and supplemental element that can be dynamically incorporated into the message construction only when the relational meaning is absent, no motor effects on constrained utterances with fixed relational meaning should be observed. On the other hand, if motion is a vital or necessary component of the production mechanisms, then motion should still influence the sentence production processes, regardless of the presence of relational meaning. Investigating both unconstrained and constrained meaning constructions will help us to precisely comprehend the underlying system of producing unconstrained or fixed messages in relation to motor activities.
CHAPTER 6
SIMULATION IN FIXED MESSAGE PRODUCTION

In this chapter, I describe a set of four experiments that explore the relationship of simulation to fixed message production. Chapter 5 demonstrated that speakers were generally influenced by motor actions when constructing a message whose internal structure was under-determined. In other words, motion is interactively utilized to determine the relational meaning of event structure, regardless of differences in the timing of enactment in relation to message formulation. These general motion effects on message construction were observed across languages. However, the results of the experiments described in Chapter 5 lead to two possible hypotheses regarding the nature of the production mechanisms. First, it may be the case that motor information is utilized as a source for configuring the event structure of the message because the production mechanism is essentially and unconditionally sensitive to the speakers’ physical activities. This would further predict that motion influences and interacts with the production mechanisms even when a relational meaning has already been associated with the message. Examining motion effects on fully established messages will address two questions regarding the role of motor information and timing when relational information is set, which were not answered by the experiments described in Chapter 5. These questions are (a) does motor information affect the conceptual components (e.g., the relational meaning) or the subsequent linguistic components (e.g., word order arrangement) in the production processes? and (b) do the motor effects change based on
the timing of when motor information becomes available in relation to the established message?

The second possibility is that motor information is incorporated into message construction only because the message is not yet associated with any particular direction-specific event structure. If this is the case, motion should not influence the constrained message production because motor information is no longer necessary or useful to supply specified directional information. In order to delineate the nature of language production mechanisms in relation to motor actions, the experiments presented in this chapter employ fixed messages whose relational structures of the event are clearly determined and established as the target conceptual forms, instead of messages whose internal structures are unconstrained. The elicited productions prompted by a sequence of pictures in this chapter are different from spontaneous speech in two main ways. First, the elicited message is intentionally constrained, while spontaneous speech is freely produced with no overt constraints. Second, in elicited productions, speakers are led to produce utterances without any communicative purposes, while the presumed goal of spontaneous speech in daily life is communication in which speakers convey their message to others. Nevertheless, these experiments are useful to tease apart the functional role of motion in producing language, that is, motion that conditionally serves as a supplement to fill in the relational meaning or motion that unconditionally serves as an influential factor in producing language.

Experiments 5–8 were conducted to explore the question of whether or not motion conditionally or unconditionally interacts with language production processes. If motor actions influence the process of producing a fully determined message, such interactive
relations between motion and speech may cause facilitative or hindering effects on speech articulation, which may be manifested by either speeding up or slowing down the onset of utterances. This would support the first hypothesis, that is, that the production mechanisms are constantly integrating information on speakers’ current physical status regardless of the presence of relational event structure in the message. On the other hand, if external motions have no effect on constrained-event description processes, then this would provide support for the second hypothesis, i.e., that motion plays an influential role in speech only when messages have no relational meaning. In this specific case, motion would appropriately frame the event and assign participants to appropriate roles. In other words, if the second hypothesis is correct, motion will not affect sentence production when participants’ roles are overtly established, as in fixed messages.

The role of motor action in producing language describing fixed messages is examined when speakers engage in motion before event perception in Experiments 5 (English) and 6 (Japanese), as well as when they do so after event perception in Experiments 7 (English) and 8 (Japanese). The reaction times (RTs) for speech onset and mouse movement in both languages and the word order in Japanese are analyzed as indicators of interactions among motion, motor simulation, and speech production.

6.1 Motion before formulating a fixed message

6.1.1 Participants

Fifty-four native speakers of English and 36 native speakers of Japanese participated in Experiments 5 and 6, respectively. Participants were compensated with a small amount of money for a 20-minute long experiment session, or they received course
credit in an introductory linguistics course. All participants reported normal hearing and vision.

6.1.2 Materials (English, Japanese)

Critical Photographs. The researcher created 30 critical pairs of pictures for this experiment. One picture in each pair depicted the initial state of an event, while the other showed the final state of the same event. Taken together, each pair represented one continuous action involving either a toward- or away-motion. Such a sequence of static photographs representing initial and final conditions of objects in an event was expected to evoke mental simulation based on two previous studies. Freyd (1983) discovered that when people perceived a sequence of photographs depicting a single event at two different temporal points, they mentally simulate (or internally experience) an implied dynamic motion. Kourtzi and Kanwisher (2000) conducted imaging studies that demonstrated that people processed implied dynamic motor information even when perceiving a single static photograph depicting an object in motion.

When the pictures were prepared, the shooting angle was maximally controlled so that the perspective or view point on the event is consistently maintained throughout all items. Each event was carefully selected so that reversing the order of the first and second pictures in critical items would produce a depiction of the same event with an opposite motor direction. For example, a picture of a hand grabbing an apple right above a plate followed by another picture of a hand holding an apple away from a plate could be interpreted as depicting the toward-action of taking the apple from the plate. Reversing
the order of those pictures could be interpreted as depicting the away-action of *putting the apple on the plate* (See sample pictures in Table 6-1).

Table 6-1: A sample pair of pictures that depict toward- or away-events

<table>
<thead>
<tr>
<th>Event types</th>
<th>Picture 1</th>
<th>Picture 2</th>
<th>Sample event descriptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Toward-event</td>
<td><img src="image1" alt="Sample Picture" /></td>
<td><img src="image2" alt="Sample Picture" /></td>
<td>Taking the apple from the plate</td>
</tr>
<tr>
<td>Away-event</td>
<td><img src="image3" alt="Sample Picture" /></td>
<td><img src="image4" alt="Sample Picture" /></td>
<td>Putting the apple on the plate</td>
</tr>
</tbody>
</table>

All together, therefore, this experiment contained 60 critical items (30 toward-events and 30 away-events; see Appendix F for a complete list). Clear event directionality distinguishes the experiments presented in this chapter from those in the previous chapter, where event directionality was flexible rather than established.

**Filler Photographs.** Thirty pairs of filler items were also created. They depicted motion events that elicited descriptions involving various action verbs (e.g., *hammering*, *ironing*, or *sharpening*) that were not likely to be used in describing toward- or away-motion in any of the critical pictures. This aimed at preventing lexical priming effects as much as possible. The more verb variations speakers produce, the more confidently we can state that their utterances are affected by motion rather than by lexical priming due to the
frequent use of particular lexical items. Filler pictures depicted various common events such as *peeling a banana, folding a napkin, hanging up the phone, or cutting a kiwi in half*. (See Appendix G for a complete list of the 30 filler picture items).

*Picture Norming (Sentence continuation task).* A norming study with four native speakers of English and 14 native speakers of Japanese who did not participate in the main experiment ensured that the 60 pairs of critical pictures clearly evoked direction-specific actions and that the 30 filler events evoked their intended, non-direction-specific actions. Participants sat in front of a computer in a quiet location and watched animated slideshows created in Power Point. The first picture depicted an initial point of an action and appeared on the screen for 800 ms, followed by the second picture, which depicted the final state of that action for 800 ms. Half of the participants (i.e., two English and seven Japanese speakers) received 30 critical events (i.e., 15 events with toward-motion and another 15 events with away-motion) and 30 filler items. The other two English speakers and seven Japanese speakers saw the counterparts of these 30 critical events (i.e., the first 15 events with away-motion and the last 15 with toward-motion) and the same 30 filler events.

Participants received papers on which was written the sentence fragment “I am” in their native language. In this sentence continuation task, they were instructed to complete the “I am” sentence by writing a short description of the dynamic event they had just seen on the screen. Items that were described similarly by more than 80% of the 18 participants were included as experimental items. Since the items were temporally dynamic, they clearly conveyed their intended actions; participants gave generally
consistent responses across all items, although some linguistic variations were observed (e.g., \textit{taking off the skin of the banana} vs. \textit{peeling the banana}). Therefore, no items were excluded.

6.1.3 Procedure

Experimental procedures were identical to those in Experiments 1 and 2, described in Chapter 5 (See section 5.1.3 for the detailed procedures), except that, as explained above, the event directionality was established rather than flexible. Participants completed a set of eight practice trials to ensure their understanding and to familiarize them with the task. Each participant was then presented with the 30 critical trials, randomly interspersed with the 30 filler trials. The 30 critical items were divided into 15 items with toward- and 15 with away-events/language, and each direction type was further divided into three groups: five after toward-motion, five after away-motion, and five after left- or right-motion.

Following the researcher’s instructions, participants positioned the mouse in the center of the clock-patterned mouse pad. For each trial, participants saw a cross for 500 ms followed by one of the four numbers (3, 6, 9, or 12) on the mouse pad. They were instructed to move the mouse toward the matching number on the mouse pad as quickly as possible. When participants left-clicked the mouse upon reaching the number, the number screen was replaced by the next screen displaying the text “I am…,” which remained for 500 ms. The mouse was kept in that position. Immediately after the “I am…” screen, a sequence of pictures was displayed in rapid succession (300 ms for each presentation). The first picture showed the initial state of the event and the second
showed the resultant state or the end point of the event. Examples of the screen sequences are presented in Figure 6-1 for Experiment 5 in English and Figure 6-2 for Experiment 6 in Japanese. As the norming study demonstrated, the intended message was clearly extractable and the event depictions were not directionally ambiguous.

Participants were instructed to describe the depicted event by verbally completing the sentence as quickly as possible, without repeating the “I am” prompt or using hesitations (e.g., “uh,” “um”). As soon as speech was captured by E-Prime, the speaker icon appeared on the screen to notify participants that their voice was being recorded. Upon completion of the sentence, participants left-clicked, repositioned the mouse back at the center of the mouse pad, and clicked once more. As with previous experiments, the xy-coordinate positions of the two mouse clicks (i.e., the first one in the center of the pad and the second one near the number on the pad) were recorded by E-Prime. In addition, E-Prime recorded the onset timing of the first sound of participants’ speech. Each event description produced by participants was separately recorded by the voice recorder.
6.1.4 Conditions

Experiments 5–8 in this chapter employed two variables: (1) physical movements in one of four directions (i.e., toward, away, left, or right), and (2) pictures that depicted direction-specific events (i.e., toward- or away-events). Right- and left-motions were merged into one condition called the “left/right condition,” resulting in three motion priming conditions. The three levels of motor priming with the two levels of event directionality produced six total conditions. The toward- and away-motions served as critical conditions, since the experiment centered on investigating their effects on utterances containing either compatible or incompatible directional information. As a control condition, right- or left-motor effects on the toward- or away-event productions were examined, because unrelated motions should not cause facilitation or inhibition in producing toward- or away-language. The experiment was designed as a Latin square, so that each of the 60 pairs of critical pictures appeared in only one motion condition for each participant.
6.1.5 Data Analysis

**Data Coding for Language Type.** A total of 3240 responses from 54 native speakers of English and 2160 responses from 36 native speakers of Japanese were recorded in the main experiments. The same two undergraduate research assistants (one native speaker of English and one bilingual speaker of English and Japanese) who worked on Experiments 1–4 (See section 5.1.5) worked with the researcher (a native Japanese speaker) to manually transcribe the responses, individually assigning each response into one of four categories: toward, away, neutral, or other, according to criteria identical to those described in Chapter 5 (See section 5.1.4 for details). To ensure the accuracy of sentence categorization, all coders were blind to the conditions. Any disagreements between coders were resolved by discussion between them. The purpose of categorizing the language type was to verify that participants paid close attention to the sequence of pictures and were able to extract the intended events with specific directionality. If participants successfully understood the pictorial scenes, then each of them was expected to produce 60 verbal event descriptions, including 15 using toward-motion language and 15 using away-motion language (i.e., for the critical items) and 30 using various kinds of language (i.e., responses for the filler items that denoted various activities not involving toward- or away-motions). Only trials with toward- or away-language in critical responses (1524 out of 1590 English critical responses and 898 out of 956 Japanese critical responses) were further analyzed.

**Data Coding for Word Order in Japanese.** Following the same reasoning discussed in Chapter 5 (See section 5.1.5), word order in the Japanese responses was systematically
coded as canonical (i.e., Loc-DO), scrambled (DO-Loc), or other for further analysis. As in the previous experiment, participants’ choice of word order was used as an indicator of the existence of simulation in producing the fixed-event descriptions. In Experiments 5–8, in contrast to Experiments 1–4, although participants were instructed to describe the depicted event as clearly as possible, they were not obligated to use the names of all the objects that were presented in the pictures. As a result, it was natural for them to focus on the event itself and say *putting down the can* instead of *putting down the can on the tray* in Japanese. Those responses that did not involve both of the two objects, namely, an entity expressed as a direct object such as *can* and a location expressed in a locative phrase such as *on the tray*, were coded as other. Relative clause utterances (e.g., *taking the can that is placed on the tray*), ungrammatical utterances, and skipped trials were also coded as other.

**Accuracy for Mouse Movement.** As described in section 5.1.5, the two mouse-click positions—the one in the center of the pad and the subsequent one near the number on the pad after the number display—were recorded by E-Prime.

**Reaction Time for Speech Onset.** E-Prime recorded the onset time of each verbal response. This was analyzed later to examine the presence of mental simulation during speech production. The motivation for using speech onset as one of the dependent measurements is explained in section 6.1.6. As in the previous experiments, this measurement was useful for investigating the internal relationship between motion and language that describes motion events.
**Data trimming.** In order to meet the assumption of parametric tests (e.g., repeated measures ANOVAs), the normality of the RT data (including speech onset and mouse movement) in Experiments 5–8 is evaluated by using a Kolmogorov-Smirnov test (KS-test). If the data are not normally distributed due to some long RTs in the data set, reciprocal data transformation is applied to reduce the impact of the large scores, and then the normality is tested again. In some individual cases, non-normality is still found (details are described in the sections corresponding to those individual cases). In future analyses, I plan to apply mixed-effect logistic regression to take a closer look at these particular cases.

**Selected Participants.** Due to equipment failure, the data from one English-speaking participant were not recorded into E-Prime, resulting in a total of 53 participants, that is, unequal numbers of participants for each language. To maintain the balanced experimental design, five additional participants were eliminated and further analyses were conducted on the remaining 48 participants. First, the accuracy of participants’ mouse movements was examined in order to eliminate any whose accuracy was below 70%. No participants were eliminated either from the English (the average accuracy was 91%) or from the Japanese experiment (the average accuracy was 89%). Second, any responses that were inconsistent with the intended message were omitted (including ungrammatical responses or neutral responses denoting no direction-specific motion), resulting in less than 4% of the English data (55 out of 1440 responses) and less than 6% of the Japanese data (58 out of 956 critical responses) being removed. No lexical persistence was observed in either language. Also, no participants in either language had
trouble understanding the experimental task, and all trials appeared to be smoothly conducted. Finally, obvious outliers (below 100 ms and above 4000 ms) were eliminated from the English data. Since the KS normality test showed that the data were not normal due to some longer RTs, the data were normalized by applying an inverse transformation. As for the Japanese data, obvious invalid outliers (e.g., above 10000 ms) and responses whose speech onset times were above or below the 2.5 SD of the subject and item means were omitted. These eliminations due to RT analysis represented 6.8% (61 out of 898 responses) of the Japanese data. As a result, 48 English speakers (i.e., eight participants each for six conditions) and 36 Japanese speakers (i.e., six participants each for six conditions) were selected for further analyses.

6.1.6 Predictions

Predictions for Language Content. Because the picture sequences clearly conveyed the intended events (confirmed by the norming study described in section 6.1.2), the meaning of participants’ responses should be fairly consistent across participants, although synonymic expressions could be produced (e.g., *put* and *placed* or *take out* and *take off*).

Predictions for Speech Onset. If language processors are sensitive to the speakers’ previous physical activities even when the details of the message have been already determined (i.e., the internal structure of the message is established so that it has little, if any, linguistic flexibility) and such activities have robust effects on language production processes, then this sensitivity should be measurably reflected by how quickly or slowly speakers articulate the first sound in event descriptions. This assumption produces two
different hypotheses regarding the facilitative or inhibitory motor effects of the compatible event type. First, if language processors are sensitive to the different degrees of event specificities that are activated by manual motion (i.e., general experiences that involve directional information) and picture perception (i.e., detailed motor information that is associated with a specific event), then the non-integratability of the two events will cause difficulty and produce delays of speech production (i.e., match inhibition). Second, if language processors are more sensitive to the general features or general conceptual domain of toward- or away-motor experiences (while ignoring the subtle differences), then compatible directional information between motor action and picture perception will result in an increase of the speed at which participants articulate the event descriptions (i.e., match facilitation).

On the other hand, if language processors are not susceptible to the speakers’ motor information when the relational meaning has been selected for the message, then no difference in speech initiation times should result from the motion types.

Predictions for Word Order in Japanese. The structural flexibilities of Japanese enable us to utilize word order (in addition to speech onset) as a measurement to indicate whether or not previous body movements influence the subsequent process of fixed-message production. As discussed in the previous chapter, speakers are more likely to produce utterances in canonical word order (i.e., Loc-DO) when describing toward-motion events because this word order reflects a consistent temporal order within the simulated scene for these stimuli. For the same reason, more utterances with scrambled word order (i.e., DO-Loc) will be produced in describing away-motion events.
6.1.7 Results

Table 6-2 summarizes the English (Experiment 5) and Japanese (Experiment 6) speakers’ reaction times (elapsed time between end of picture stimulus and onset of speech) for critical items. The average mean reaction times are presented in terms of three motions (away-, toward-, or left/right-motions), and sentence types and standard deviations are shown in parentheses in the table. Since both left- and right-motion conditions serve as indicators of the baseline speech onset times for Experiments 5 and 6, they are merged into the left/right-motion condition.

Table 6-2: RTs of speech onset with away- or toward-responses in motion-first constrained message-second design:

<table>
<thead>
<tr>
<th>Picture/Event Type</th>
<th>Experiment 5 (English)</th>
<th>Experiment 6 (Japanese)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Away Motion</td>
<td>Toward Motion</td>
</tr>
<tr>
<td>Away Lg.</td>
<td>1294 (753)</td>
<td>1405 (805)</td>
</tr>
<tr>
<td>Toward Lg.</td>
<td>1210 (608)</td>
<td>1255 (601)</td>
</tr>
</tbody>
</table>

6.1.7.1 English speakers (Experiment 5)

The relationship between two factors, that is, picture or event types (i.e., toward-vs. away-events) and motion compatibility (i.e., “match motion” whose directionality is compatible with the perceived picture, “neutral (left/right) motion” whose directionality is unrelated to pictures, and “mismatch motion” whose directionality is opposed to the depicted event) was statistically examined in the two-way repeated measure ANOVAs. The analyses revealed no main effect of picture type/event direction in the subject analysis (F₁(1,47)=.6, p=.44, η²_p=.01) or in the item analysis (F₂(1,29)=.23, p=0.63, η²_p=.008). Contrary to the predictions, motion produced no significant main effect (F₁
(2,94)=.22, \( p=0.81, \eta^2_p=0.005 \), \( F_2 (2,58)=.92, \ p=0.4, \eta^2_p= 0.03 \), which means that previous motor action did not influence the speed of producing descriptions of pictured events. As shown in Figure 6-3, no interaction effect between picture/event type and motion was observed (\( F_1 (2,94)=0.9, \ p=.41, \eta^2_p=.02 \), \( F_2 (2,58)=1.0, \ p=.37, \eta^2_p=.03 \)).

![Figure 6-3: Speech initiation times (milliseconds) of toward- and away-language after match, neutral, or mismatch motion in Experiment 5](image)

Figure 6-3: Speech initiation times (milliseconds) of toward- and away-language after match, neutral, or mismatch motion in Experiment 5

In addition, two one-way repeated measures ANOVAs were conducted on away-event and toward-event descriptions to take a closer look at each language type; in both analyses, speech onset times were dependent measures while the three levels of motion type (match, mismatch, or neutral motions) were independent measures. No motor effects on utterance initiations of toward- or away-event descriptions were found in either participant or item analyses (toward-language: \( F_1(2, 94)=.31, \ p=.73; \ F_2(2, 58)=.34, \ p=.71 \), away-language: \( F_1(2, 94)=.8, \ p=.45; \ F_2 (2, 58)=2.0, \ p=.15 \)). Paired t-tests confirmed that the numerical difference between toward- and away-language in the mismatch condition was not significant (\( t(47)=1.5, \ p=.13 \)).
6.1.7.2 Japanese speakers (Experiment 6)

The same sets of analyses that were used for the English data were also conducted with the Japanese data. Figure 6-4 illustrates the speech initiation times in milliseconds. Two-way repeated measure ANOVAs showed no main effect of picture type/event direction in the subject analysis ($F_1(1,35)=2.0$, $p=.2$, $\eta^2_p=.05$) or in the item analysis ($F_2(1,29)=3.0$, $p=.09$, $\eta^2_p=.09$). Critically, physical motion revealed a significant main effect in the subject analysis ($F_1(2,70)=9.7$, $p<.001$, $\eta^2_p=.22$) and in the item analysis ($F_2(2,58)=4.1$, $p<.05$, $\eta^2_p=.13$), which indicates that speech initiating times were significantly different depending on whether motion direction is compatible, incompatible, or neutral with the event direction described in the utterances. No interaction between motion and language type suggests that motion influenced the speech initiation times in similar ways, regardless of the event types ($F_1(2,70)=.25$, $p=.78$, $\eta^2_p=.007$; $F_2(2,58)=6.7$, $p=.52$, $\eta^2_p=.02$).

![Figure 6-4: Speech initiation times (milliseconds) of toward- and away-language after match, neutral, or mismatch motion in Experiment 6](image-url)
In addition, speech onset times of toward and away responses were separately analyzed in terms of subjects and items. First, one-way repeated measures ANOVAs with speech onset times of away responses as the dependent measure and three levels of motion prime type as the independent measures showed significant effects in both subject and item analyses ($F_1(2, 70)=5.7, p<.01; F_2(2, 58)=3.7, p<.05$). The subsequent paired t-tests on away responses showed a significant match inhibitory effect, that is, away responses were produced more slowly when the direction of the motor actions matched the direction conveyed by the subsequent speech than when the direction of the action was unrelated to the message or neutral (i.e., left or right condition) ($t_1=3.0, p<.01; t_2=2.8, p<.01$). Although the speech initiation of away responses also appears to be slower after toward-motion than after neutral motions, the difference was not significant ($t_1=1.7, p=1.1; t_2=1.6, p=.1$). The speech initiation times for away-language were significantly different after match (away) motion and after mismatch (toward) motion in the subject analysis ($t_1=2.0, p<.05; t_2=.97, p=.34$).

Second, one-way repeated measures ANOVAs with toward responses similarly revealed significant effects in subject analyses ($F_1(2, 70)=3.7, p<.05$), but did not reach the significant level by item analysis ($F_2(2, 58)=1.8, p=.18$). Paired t-tests showed a match inhibition in toward responses, that is, the initiating times of toward responses were slower after executing toward-motion than after executing unrelated left/right-motion ($t_1=2.1, p<.05; t_2=1.6, p=.12$, although it was only significant by the subject analysis). Toward responses in the mismatch motion condition were slower than those in the neutral motion condition in the subject analysis ($t_1=2.2, p<.05$), but this did not reach
significance in the item analysis ($t_2=1.6$, $p=.13$). Plausible implication derived from this general tendency of match/mismatch inhibition in toward- and away-language will be discussed in the following section.

Whether or not word order might also reflect the temporal aspects or orders of the event being simulated was examined in toward- and away-language. As mentioned above, less than 6% of the utterances did not involve both DO and locative phrases and were categorized as other. Certain items in particular tended to be described in other-type utterances. For example, one of the critical items was *coffee can*. Four percent of the utterances for this item used canonical word order (e.g., ‘from-the-coffee-can lid-Acc taking’), 4% were in scrambled word order (e.g., ‘lid-Acc on-the-coffee-can putting’), and 92% were in the other category (e.g., ‘lid-Acc closing’ or ‘lid-Acc opening’). If items were not generally described by using both DO and Loc, but rather described with other language for more than 70% of the data, they were excluded for the purpose of the word order analysis. This left 20 items out of 30 items for the word order analysis.

As discussed in Chapter 5, examining word order can illuminate two intriguing questions about word order determination in relation to event construction and motor activity: (a) are speakers more likely to arrange the word order according to the temporal order of the simulated event? and (b) in which motion circumstance/condition is such temporal consistency between event and word order most likely to be observed?

In order to investigate the first question, percentages of word order preferences in toward- or away-language were calculated in the same way as discussed in section 5.1.7.2. Two paired t-tests were conducted on these 20 items for each (toward and away) language type. As shown in Table 6-3, the t-tests revealed that toward-language
utterances were more likely to have canonical word order than non-canonical word order ($t_1=12.3, p<.001; t_2= 7.1, p<.001$), while away-language utterances were expressed more often in scrambled word order ($t_1=13.4, p<.001; t_2= 12.3, p<.001$). These results, which suggest that internal/mental experiences of a simulated event are more likely to arrange the word order chronologically, raise the second question, i.e., whether or not motion is the causal source that inspires a message with a compatible directionality and that sequentially determines word order.

Table 6-3: Number of responses with canonical word order, scrambled word order, or other word order in Japanese in motion-first, constrained-message-second design

<table>
<thead>
<tr>
<th>Sentence Type</th>
<th>Away Motion</th>
<th>Toward Motion</th>
<th>Left/Right Motion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Away Lg. (Canon.)</td>
<td>6</td>
<td>5</td>
<td>3.5</td>
</tr>
<tr>
<td>Away Lg. (Scram.)</td>
<td>94</td>
<td>71</td>
<td>36</td>
</tr>
<tr>
<td>Away Lg. (Other)</td>
<td>59</td>
<td>49</td>
<td>28.5</td>
</tr>
<tr>
<td>Toward Lg. (Canon.)</td>
<td>46</td>
<td>34</td>
<td>22</td>
</tr>
<tr>
<td>Toward Lg. (Scram.)</td>
<td>49</td>
<td>19</td>
<td>20</td>
</tr>
<tr>
<td>Toward Lg. (Other)</td>
<td>71</td>
<td>60</td>
<td>33.5</td>
</tr>
</tbody>
</table>

As Figure 6-5 illustrates, with toward-language, the percentage of canonical word order utterances is the highest after toward-motion (i.e., 64%) compared to after left/right- (i.e., 52%) or away-motion (i.e., 48%). Likewise, with away-language, the percentage of scrambled word order utterances is higher after away-motion (i.e., 94%) than after left/right- (i.e., 91%) or toward-motion (i.e., 93%), although the difference is much smaller. This contributes evidence that word order is influenced by event type as well as by motor activities that speakers have engaged in.
Figure 6-5: Counts and percentages of toward and away responses with canonical or scrambled word order after performing motion in specific directions (Experiment 6)

6.1.8 Discussion

The results from Experiment 6 in Japanese suggest that motions that speakers have previously executed influence how quickly they can start describing a particular type of dynamic event. Away-event descriptions as well as toward-event descriptions are produced significantly slower when they match the previous motion (i.e., away-motion or toward-motion, respectively) than when the previous motion is unrelated (i.e., left/right-motion). This match inhibition effect seems to be weaker in toward-event descriptions than in away-language. More crucially, inhibitory effects are also found in away- and toward-language when motion and event type mismatch.

These results show the characteristic that illustrate the interactive nature of production system, that is, in general, motion significantly affects sentence production, although the relational meaning is fully established in the message. Moreover, these results suggest two factors that may explain the various facilitative or inhibitory motor
effects on utterances. The first plausible factor in these match/mismatch inhibitions could be Bergen et al.’s (2003) claim that related but nonidentical neural activations hinder the semantic processes in sentence comprehension. Expanding their claim to this production study, match and mismatch inhibitions may have been observed because the two pieces of motor information (that are derived from mouse movements and the depicted event) activate related but not exactly identical neural resources (due to the shared global/general toward- or away-directionality). Toward-motion and away-motion interact with the event message in similar ways because they belong to the same vertical motion domain.

Second, a plausible factor that could be responsible for the stronger motion-message inhibition in the match condition than in the mismatch condition is integratability, which refers to the question of whether motor information activated by physical motion can be integrated into the motor simulation evoked by event perceptions (Kaschak et al., 2005). More specifically, general motor information activated by manual movements cannot be entirely integrated into a simulation that involves event-specific motion because the motions are significantly different in terms of level of action specificity (e.g., schematic knowledge of toward- or away-motion vs. detailed hand shapes or particular angles of arm movements. As a result, two factors—recruitment of related but not identical neural activations and non-integratability due to the different motor specificities—may explain why the speed of producing utterances is slower in the match condition than in the neutral condition.

In addition, word order analyses in Japanese show that speakers selectively choose a word order that is consistent with a particular temporal sequence of toward- or
away-events. This indicates that because speakers mentally simulate a direction-specific event when creating a message, they tend to arrange the word order as the event unfolds. Furthermore, this tendency toward “message–word order match” becomes most significant when speakers perform the compatible motion prior to the utterance formulation. This suggests that the general knowledge of directionality (i.e., the temporal order of mental access is location-object in toward-events and object-location/destination in away-events) that is activated by physical motion affects the word order alignment. Although manual knowledge activated by motion can not be completely integrated into the event-specific motor information (such as hand shapes or arm angles), temporal knowledge of toward- or away-events is activated by motion and influences the process of arranging word order.

It may be due to the different linguistic properties of the two languages that the English speakers in Experiment 5 were not similarly influenced by their previous body experiences. However, this needs further investigation.

6.2 Motion while formulating a constrained message

Experiments 7 and 8 in this section are identical to Experiments 5 and 6, except that the order of the physical motions and picture stimuli are reversed. Participants first perceive the “I am…” screen followed by a sequence of two pictures; afterwards, they move the mouse in the specified direction, and then describe the depicted scene in a simple sentence (See detailed procedures in section 6.2.3). Therefore, in this design, motor action is extremely cognitively salient, as it takes place immediately before
articulation. This allows us to examine the immediate motor effects on the process of producing the constrained/fixed messages.

6.2.1 Participants (English, Japanese)

Sixty native English speakers and 36 native Japanese speakers participated in the experiments, and each was compensated with a small amount of money or they received course credit for a 20-minute experiment session. All English speakers were recruited from the University of Hawai‘i at Mānoa while all Japanese speakers were from Hiroshima University.

6.2.2 Materials (English, Japanese)

The picture materials for Experiments 7 (English) and 8 (Japanese) were identical to those used in Experiments 5 (English) and 6 (Japanese).

6.2.3 Procedure

Procedures were identical to those of Experiments 5 and 6, with the one exception already stated: in Experiments 7 and 8, participants saw the pictures depicting dynamic scenes before they moved their hand rather than afterwards. Examples of sequences of screens are shown in Figure 6-6 for Experiment 7 and Figure 6-7 for Experiment 8. First, the screen with the sentence fragment “I am…” appeared for 500 ms, followed by two pictures in rapid succession (300 ms for each representation), depicting an initial and a final state of a continuous motion. As in the previous experiments, participants then saw a cross on the center of the screen for 500 ms, followed by one of the four numbers 3, 6, 9,
or 12. As soon as participants moved the mouse near the given number on the mouse pad
and left-clicked, the screen disappeared. Participants continued a sentence starting with “I
am,” which described the perceived scene (e.g., *putting the apple on the plate*) as quickly
as possible, without verbally repeating “I am.” The verbal responses triggered the screen
with a speaker icon, and the E-Prime program recorded the response times of the speech
onset. Following sentence completion, participants clicked the mouse near the clock
number, repositioned it back to the center of the mouse pad, and clicked again.

Figure 6-6: Experiment 7 (English). Away-message followed by left-motion

Figure 6-7: Experiment 8 (Japanese). Toward-message followed by right-motion
6.2.4 Conditions

Six conditions identical to those in Experiments 5 and 6 were created (i.e., two depicted event directions, toward and away, and three hand motions, toward, away, and right/left).

6.2.5 Data Analysis

Data Coding for Language Type. The same two undergraduate research assistants (one native speaker of English and one bilingual speaker of English and Japanese), who worked on the previous experiments, worked with the researcher (a native Japanese speaker) to manually transcribe a total of 3600 verbal responses from 60 English-speaking participants and 2160 from 36 Japanese-speaking participants in the main experiments. Then, each coder independently categorized the critical responses (i.e., 1800 responses in English and 1080 responses in Japanese) into one of the four sentence types (i.e., toward-, away-, or neutral-sentence, or other) according to the coding criteria for language types established in Chapter 5. These sentence categorizations verify whether participants correctly understood the experiment and reliably paid attention to each trial. If they did, then each participant was expected to produce 15 toward- and 15 away-direction statements for critical items and another 30 various event descriptions for filler items.

Data Coding for Word Order. The Japanese-speaking researchers (one research assistant and the researcher) also labeled the word order of all critical Japanese responses as one of
three types: canonical (Loc-DO), scrambled (DO-Loc), or other (i.e., ungrammatical sentences, neutral sentences involving non-directional verbs such as *watching*, *heating*, or *baking*, or sentences without both a locative phrase and a direct object; for example, relative clause sentences such as *lifting the rack which can hold at least ten basketballs*). Any disagreements were discussed and resolved between coders. This word order labeling was important for investigating the relationship between word order variations and event types.

**Accuracy and Response Time for Mouse Movements.** The two mouse-click positions—in the center of the pad and subsequently near the number on the pad after the number display—were recorded by E-Prime to ensure participants’ accuracy of hand motion. In addition, the response times showing how quickly participants could make that move were recorded.

**Speech Onset.** The onset of speech production was recorded by E-Prime. It was then analyzed to assess whether compatible or incompatible directionality between mouse motion and the linguistically motivated mental simulation generated by event descriptions would be reflected by speech initiation times.

**Selected Participants.** The validity of participants was examined based on several criteria, including their accuracy in describing the depicted events, the accuracy of their mouse movements, the exhibition of lexical persistency, and their apparent understanding of the experimental task. No English or Japanese participants were excluded based on the
first criterion; each participant correctly described the depicted events over 95% of the time. No English or Japanese participants were eliminated due to mouse inaccuracy since the average accuracy was above 70% (89.3% in English and 85.6% in Japanese). No indication of lexical persistence or difficulty in understanding the experimental task was observed in either experiment. Data from an equal number of participants for each condition within each language (60 English speakers, or 10 participants for each of six conditions, and 36 Japanese speakers, or six participants for each of six conditions) were further analyzed to investigate the relationship between motion and language production mechanisms.

### 6.2.6 Predictions

**Predictions for RTs for Mouse Clicks.** In Experiments 7 and 8, response times for mouse clicks are the critical dependent measure to verify whether participants run the motor simulation when perceiving a sequence of initial and final states of event-depicting pictures. Previous comprehension research has shown that simulation induced by understanding sentences describing toward- or away-motion influences subsequent key-press actions involving either toward- or away-motion (Glenberg & Kaschak, 2002; See section 2.3.4 for detailed discussion). Thus, if mental simulation involving toward- or away-motion is evoked by picture perception in the experiments presented in this dissertation, then the subsequent mouse-click action (i.e., toward- or away-motion) should be hindered or facilitated depending on the directional compatibility between simulation and motion.
For instance, participants who perceive and so simulate toward-motion events were predicted to show faster response times with toward-motion mouse-clicks than with away-motion mouse-clicks. Likewise, away-simulation will facilitate subsequent away-motion mouse-clicks, evidenced by faster response times for away-clicks than for toward-clicks. This match advantage is expected because simulation generated by picture perception creates a particular activation network that is responsible for a direction-specific motion. Engaging in the subsequent compatible motion reactivates the same neural pattern, resulting in motor facilitation.

**Predictions for Language Content.** As the norming study discussed in Section 6.1.2 ensured, participants will produce event descriptions that clearly convey the implied motion in a particular direction.

**Predictions for Speech Onset.** At the point where physical motions are executed, participants have extracted the implied event information from the pictures and may have been already engaged in encoding the message with appropriate linguistic components. The predicted effects of motor activities on speech onset hinge upon two different hypotheses. First, if the production system is unconditionally sensitive to the speakers’ physical status regardless of at which point physical motions are executed in respect to the linguistic planning process, then compatibility or incompatibility between the directions denoted by messages and motions will consistently produce facilitation or inhibition in the response times for speech onset. For example, unconditionally sensitive speakers would initiate the first sound of a toward-event description faster if they move
their hand toward their body at any point just before, during, or just after the linguistic planning process than they would after moving their hand away from their body or in unrelated (i.e., right or left) directions. In the same way, speakers will start producing a sentence describing an away-motion faster immediately after moving their hand away from their body than they would after moving it toward their body or in unrelated directions. Second, if the production system is not sensitive to external information such as motion once the event structure of the message has been constructed and associated with particular relational meaning, then physical motion that occurs after message construction will not produce any RT differences in the onset of speech describing different direction-specific events.

**Predictions for Word Order in Japanese.** Following the same logic explained in Experiments 2, 4, and 6 regarding the ability of the structural flexibility of Japanese to reflect mental simulation, I predict that more sentences with scrambled word order will be produced to describe away-events while more sentences with canonical word order will be used to describe toward-events.

**6.2.7 Results**

Table 6-4 represents mouse-click response times, that is, the elapsed times measured between the first click in the center of the mouse pad, which initiates the trial, and the second click near the designated number (3, 6, 9, or 12). Table 6-5 shows response times for speech onset with away or toward responses after primed motion in English (Experiment 7) and Japanese (Experiment 8).
Table 6-4: RTs for mouse click with Away or Toward event descriptions

<table>
<thead>
<tr>
<th>Picture/Event Type</th>
<th>Experiment 7 (English)</th>
<th>Experiment 8 (Japanese)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Away Click</td>
<td>Toward Click</td>
</tr>
<tr>
<td>Away Lg.</td>
<td>2336 (614)</td>
<td>2312 (697)</td>
</tr>
<tr>
<td>Toward Lg.</td>
<td>2244 (622)</td>
<td>2275 (599)</td>
</tr>
</tbody>
</table>

Table 6-5: RTs for speech onset with Away or Toward event descriptions

<table>
<thead>
<tr>
<th>Picture/Event Type</th>
<th>Experiment 7 (English)</th>
<th>Experiment 8 (Japanese)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Away Motion</td>
<td>Toward Motion</td>
</tr>
<tr>
<td>Away Lg.</td>
<td>708 (281)</td>
<td>825 (445)</td>
</tr>
<tr>
<td>Toward Lg.</td>
<td>724 (367)</td>
<td>728 (221)</td>
</tr>
</tbody>
</table>

6.2.7.1 English speakers (Experiment 7)

First, two one-way repeated measures ANOVAs were conducted on the mouse-click response times to examine whether or not participants were engaged in motor simulation generated by picture perception. The first one was conducted on trials involving away-event descriptions and the second one on toward-event descriptions, both with response times for click actions as dependent measures and the three levels of mouse movements (i.e., toward, away, or left/right) as independent measures. As opposed to my predictions based on Glenberg and Kaschak’s (2002) study, both participant and item analyses of away- or toward-language revealed that the speed of hand motion was not influenced by direction specific simulation evoked by picture perception (away-language: $F_1(2, 118)=.41$, $p=.66$, $F_2(2, 58)=.15$, $p=.87$; toward-language: $F_1(2, 118)=.04$, $p=.96$, $F_2(2, 58)=.57$, $p=.57$).
Second, a two-way repeated measures ANOVAs was conducted to investigate the relationship between picture/event type (i.e., toward- or away-event) and motion (i.e., match, neutral, or mismatch motion). It revealed no main effect of motion in the subject analysis ($F_1(2,118)=2.3$, $p=.1$, $\eta^2_p=.04$) or in the item analysis ($F_2 (2,58)=2.1$, $p=.13$, $\eta^2_p=.07$). Main effect of picture/event type was observed not in the subject analysis ($F_1(1,59)=.56$, $p=.46$, $\eta^2_p=.009$), but in the item analysis ($F_2 (1,29)=5.2$, $p<.05$, $\eta^2_p=0.15$). This indicates a mild advantage for initiating toward-language compared to away-language in English. More importantly, a significant interaction between motion and event type was observed in the subject analysis ($F_1(2,118)=3.9$, $p<.05$, $\eta^2_p=.06$), although it was not significant in the item analysis ($F_2 (2,58)=1.4$, $p=.25$, $\eta^2_p=.05$).

The data was further sent to two one-way repeated measures ANOVAs on the speech onset times, employing the hand motions as the independent measures and the response times of speech onset in describing different types of events as the dependent measures. The first analysis conducted on away-language showed a marginal significant effect of motion in the subject analysis, but reached a significant level in the item analysis ($F_1(2, 118) =2.4$, $p=.09$, $F_2(2, 58) =3.8$, $p<.05$). To delineate the relationship between physical motion and event depiction, paired t-tests were conducted. As seen in Figure 6-8 below, participants articulated utterances describing away-events significantly faster after performing compatible away-motions than after unrelated (left or right) motions ($t_1=2.8$, $p<.01$; $t_2=2.6$, $p<.05$) (i.e., match facilitation).

The second one-way repeated measures ANOVAs was conducted on speech onset times on toward-language, employing identical dependent and independent measures. It revealed a significant motor effect on speech initiating times only in the subject analysis.
(F₁(2, 118) = 4.2, p<.05, F₂(2, 58) =.23, p =.79), and subsequent paired t-tests showed that speakers initiated utterances describing toward-events marginally faster after mismatch motions than neutral motions (t₁=1.95, p=.056; t₂=.6, p=.57).

Figure 6-8: Speech initiation times (milliseconds) of toward- and away-language after match, neutral, or mismatch motion in Experiment 7

6.2.7.2 Japanese speakers (Experiment 8)

The same analyses as were used for the English experiment were conducted with the Japanese mouse-click data. The results were similar to the English results. One-way repeated measures ANOVAs showed that the speed of participants’ hand motions was not influenced by event representations depicted by a sequence of pictures (away-language: F₁(2, 70)=1.83, p=.17, F₂(2, 58)=.67, p=.52; toward-language: F₁(2, 70)=2.5, p=.09, F₂(2, 58)=.45, p=.64).

Second, a two-way repeated measures ANOVAs was conducted to examine how speakers’ speech initiation times were influenced by the relationship between event type and motion. It showed no main effects of event type (F₁(1,35)=.28, p=.6, η²_p=.008;
F₂(1,29)=.53, p=.47, η²_p=.018) or motion (F₁(2,70)=.45, p=.64, η²_p=.013; F₂(2,58)=.11, p=.9, η²_p=.004). It is particularly important that the significant interaction between event type and motion (F₁(2,70)=7.8, p<.005, η²_p=.18; F₂(2,58)=1.3, p=.29, η²_p=.04) observed in Experiment 8 included match facilitation for toward-language and match inhibition for away-language, that is, the reverse of the effects observed in Experiment 7 (i.e., match inhibition for toward- and match facilitation for away-language).

Another set of one-way repeated measures ANOVAs conducted on toward- and away-language provided more specific characteristics of each type of language. Overall, no motor effects on speech onset were observed (away-language: F₁(2, 70)=2.1, p=.13, F₂(2, 58)=.63, p=.54). However, as shown in Figure 6-9 below, paired t-tests conducted on toward-language showed that participants initiated toward-event descriptions faster when the directions of the pictured event and the subsequent hand motion were compatible than when they were incompatible (t₁=3.2, p<.01), or when they were unrelated, although significance was only marginal (t₁=1.7, p=.1; t₂=.35, p=.73).

In contrast, paired t-tests conducted on away-language revealed that away-language was produced slower after match motion than after mismatch motion (t₁=2.0, p=.0051; t₂=.1.03, p=.31), although the difference in speech initiation after match motion and after unrelated motion was only numerically observed (i.e., 772.0 ms vs. 717.6 ms, respectively, t₁=1.7, p=.09; t₂=.91, p=.37).
Third, word order was examined to investigate two questions that have been asked throughout the Japanese experiments in this study: (a) is word order modulated to align with the temporal structures of an event? and (b) does motion have a causal role in determining word order?

No items were eliminated for this word order analysis because participants consistently utilized both movable and location objects in their production, and the percentages of canonical and scrambled word order were calculated as described in Chapter 5. Table 6-6 shows the number of tokens of toward and away responses that were expressed in canonical or scrambled word order. Proportions of canonical and scrambled word order were calculated as described in section 5.1.7.2 (e.g., proportions of canonical word order in toward--language are calculated by counts of canonical responses in toward-language/sum of canonical responses in toward-language/sum of canonical responses in toward- and away-language), in order to examine the relationship between event type and word order. The t-tests revealed that
toward-language was more likely to be expressed in canonical word order than in scrambled word order ($t_1 = 15.0$, $p < .001$; $t_2 = 10.3$, $p < .001$), while away-language was often produced in scrambled word order ($t_1 = 15.0$, $p < .001$; $t_2 = 10.3$, $p < .001$). These results verify that word order is more likely to be arranged according to the temporal construction of the event.

Table 6-6: Number of responses with canonical word order, scrambled word order, or other word order in Japanese in constrained-message–first, motion-second design

<table>
<thead>
<tr>
<th>Sentence Type</th>
<th>Away Motion</th>
<th>Toward Motion</th>
<th>Left/Right Motion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Away Lg. (Canon.)</td>
<td>11</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Away Lg. (Scram.)</td>
<td>95</td>
<td>84</td>
<td>47.5</td>
</tr>
<tr>
<td>Away Lg. (Other)</td>
<td>8</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Toward Lg. (Canon.)</td>
<td>49</td>
<td>39</td>
<td>19</td>
</tr>
<tr>
<td>Toward Lg. (Scram.)</td>
<td>59</td>
<td>38</td>
<td>21</td>
</tr>
<tr>
<td>Toward Lg. (Other)</td>
<td>12</td>
<td>8</td>
<td>10</td>
</tr>
</tbody>
</table>

This leads to the more intriguing question of whether or not motion, in addition to event type, influences word order arrangements. As Figure 6-10 illustrates, with toward-language, percentages of canonical word order are slightly higher after toward-motion (i.e., 50.6%) than after left/right- (i.e., 47.5%) or away-motion (i.e., 45.4%). With away-language, however, such a tendency is not observed. This absence of effect might be due to a ceiling effect, in which motion effects on the word order alignment cannot be revealed because away-language is very strongly biased to have scrambled order in general (i.e., proportions of scrambled word order after away-, toward-, or neutral motion: 89.6%, 92.2%, or 94.4%).
6.2.8 Discussion

The results from Experiments 7 and 8 provide two important findings that suggest how simulation is triggered by perception and interacts with motor actions during production of fixed messages. First, that there are no direction compatibility effects between perceived events and hand motions (based on the mouse-click RTs) in this production study indicates that mental simulation activated by event perceptions does not interact with the subsequent motor execution in an identical way as in comprehension processes. That is to say, these null effects seem to contradict the ACE effect (i.e., participants respond faster when motor responses match the content of the sentence) observed in comprehension studies employing the sensible judgment task where participants are consciously accessing conceptual representations and sentential meanings when they execute their motor responses (Glenberg & Kaschak, 2002) or the knob task where participants read segments of sentences as they are turning a knob clockwise or
counterclockwise (Taylor, Lev-Ari, & Zwaan, 2008). This may imply that motor information activated by reading sentences or perceiving pictures can be cognitively represented differently depending on the purposes of generating such motor simulation, that is, for comprehending meanings or creating messages. This speculation needs further exploration.

Second, the speech onset results in English and Japanese seem to indicate how motion can be utilized in the process of producing the constrained message. First, when speakers perceive the depicted event, perform motor actions, and then describe the depicted event, attention (simulation) order knowledge (i.e., attention order is object-location for away-events and location-object for toward-events) that became cognitively active due to physical motion seems to interact with the subsequent process of speech initiation. If the previous event simulation (induced by pictures) interacts with motion (which evokes motor information/attention order knowledge), then an interaction effect between event type and motion on speech RTs should be expected. More specifically, directional compatibility between simulated event and motion should consistently produce facilitation (or inhibition) while incompatibility may result in inhibition (or facilitation) of the speech onset times. However, such consistent interaction effect (between motion and message) was not observed in either the English or the Japanese results. Rather, the data seems to support the idea that when the detailed meaning of the event has already been fixed, motion interacts with the grammatical encoding components, not with the conceptual components.

Because all of the necessary elements or ingredients of the message are completely available in Experiments 7 and 8, there is no room (at the conceptual level)
for motor information to intervene or have an influence. As a result, attention order knowledge evoked by physical motion may interact with the subsequent grammatical process, that is, word order arrangement. The interaction between event order and word order produces two predictions regarding facilitation and inhibition effects of speech onset. First, if attention order knowledge interacts with word order arrangement, then consistent order between attention and word arrangement should predict facilitation effects, regardless of event or language types. Motor information, specifically, its attention order of the event, functions as a facilitator to align the message into a particular order in the grammatical encoding process when motion and word orders are consistent. Second, the reverse should also be true, that is, motion interferes with the word order encoding, resulting in slow speech initiation times when motion and word order are inconsistent.

In English, because the dominant word order in this study is DO-Loc, which is compatible with attention order induced by away-motion, facilitation should be consistently observed for utterances produced after away-motion, but inhibition should be observed after toward-motion. Regardless of the event type, the statistical analyses demonstrate (a) significant facilitations when motion (or more appropriately, attention order knowledge evoked by motion) and word order share the same order in accessing the DO and the location (i.e., away-motion–toward/away-language facilitations) and (b) numerically informative inhibitions when motion and word order are different (i.e., toward-motion–toward/away-language inhibitions) are indeed observed.

These same hypotheses seem to hold true in Japanese language. Canonical word order in Japanese is Loc-DO, which is consistent with the attention order generated by
toward-motion. Therefore, facilitative speech initiations should always be observed after toward-motion while difficulties in initiating utterances should be observed after away-motion regardless of the event type. In fact, as the statistical analyses show, there are (a) facilitations in initiating utterances when motion and word order activates or refers to the entities in an event in the same order (i.e., a significant facilitation in the toward-motion–toward-language condition and a numerical facilitation in the toward-motion–away-language condition), and (b) inhibitions in initiating utterances when motion and word order mismatch (i.e., a significant away-motion–away-language inhibition and a numerical indication of away-motion–toward-language inhibition).

Although these statistically weak effects in Experiments 7 and 8 do not fully support the claim that motion influences the word order encoding process, rather than conceptual production processes, they may nevertheless provide a starting point for further exploration of how and when external information intervenes in spontaneous speech production in more precise ways.

6.3 General discussion

This chapter explores the role of motor actions in producing messages whose relational meaning is already entirely determined at the time of processing. Experiments 5 (English) and 6 (Japanese) examine whether or not language production mechanisms are sensitive to the cognitive information prompted by specific motions into the subsequent process of fixed-message construction. To assess the consequence of relative timing of motion and picture perception in producing fixed messages, the order of motion and picture perception used in Experiments 5 and 6 is reversed in Experiments 7
(English) and 8 (Japanese) (i.e., in Experiments 7 and 8, participants first see the pictures, then perform the designated motion, and finally verbally describe the pictured events). Experiments 7 and 8 allow us to examine whether language processors incorporate directional information activated by physical motion into the ongoing process of fixed-message construction.

The results from Experiment 6 indicate that language processors are sensitive to previous motor information even when the relational meaning is determined in the message. In general, integratability or conceptual specificity plays an important role. That is to say, if the motor information activated by physical movements and the motor information encoded in specific events are related, but not identical (i.e., two pieces of motor knowledge share general directionality, but differ in detailed actions), they are not integratable, resulting in processing difficulties that are reflected in slower speech onset times. Integratability can be seen as conceptual specificity, that is, motion enactment generates general motor information while picture perception activates specific event-information. The absence of motion effects in English suggests a speculation regarding the persistency of motor information: it may indicate that previously activated motor information may interact with the conceptual production processes, but speech initiation times do not reflect such interaction because motor information decays and becomes too weak to make an impact on utterance initiations.

Experiments 7 and 8 show that language processors are also sensitive to concurrent motor information when the message has already been cognitively active. As discussed in the previous section, when the relational meaning has been established in the message, motion seems to interact with the subsequent grammatical encoding process,
namely, word order arrangement, instead of with the conceptual process. Across languages, facilitations of speech initiation times are generally observed when the attention order generated by specific directional motion and word order follows the same accessing order as the DO and location, while the reverse is true for inhibition effects. Moreover, word order analyses in Experiments 6 and 8 show that Japanese speakers unconsciously prefer to encode toward- or away-events in a temporally consistent word order, namely, toward-events in canonical word order and away-events in scrambled word order. More interestingly, this chronological order consistency between event type and word order gains further support from another factor, motion. When conducting toward-motion, agents initially identify the location followed by the target object, while this order is reversed when they perform away-motion. The fact that event–word order consistency becomes most significant after a corresponding (compatible) motion indicates that this consistency is not merely a correlation between event language and word order, but provides additional evidence to support the argument that general embodied knowledge (that is activated by toward- or away-motor activities) affects not only the event formulation but also word order determination.

These four experiments illuminated several general characteristics of how the role of motion in producing unconstrained messages changes according to the timing of the motion execution relative to the message formulation. First, body movements (e.g., toward- or away-motions) activate appropriate cognitive domains that store direction-specific experiences accumulated in daily life. Second, the language production mechanisms seem to be essentially responsive to and consistently interact with speakers’ cognitive status. Directional information activated by physical activities can affect
speakers’ speech initiation times for unconstrained and constrained messages. That is, there is an effect even when such motor information is no longer necessary as a source of relational determination of the message, regardless of the timing of motion enactment relative to event perception. Third, the timing of when motor information becomes available to language processors changes its functional role in the production of utterances. When motor information becomes available first, before the next process, that is, encoding event-specific information, then conceptual specificity is an influential factor that affects speech onset times, as discussed in Experiment 6. However, when the event has been completely established first, and motor information is activated second, then consistency of attention order (motion)–word order seems to be the factor that determines speech initiation times.

Based on the findings reported in this and the previous chapters, Chapter 7 proposes a production model designed to exemplify more precisely the role of physical motion in meaning generation and message encoding processes when producing both unconstrained and constrained messages.
CHAPTER 7

GENERAL DISCUSSION: ACTION-IN-MESSAGE (AIM) EFFECTS

This study has investigated the effects of physical movements on producing utterances, in English and Japanese, whose relational meanings are internally determined or under-determined. In other words, the motivation of the study was to explore the effects of action in message formulation, or action-in-message effects. These action-in-message (AIM) effects were examined in a series of experiments that manipulated the timing of motor activities relative to picture perception during the process of producing sentences. Timing was manipulated in order to assess whether such relative motion timing can vary the degree of impact (e.g., present or absent, strong or weak) or the type of interactive effects (e.g., facilitation or inhibition) that motor activities may bring to sentence production.

Chapters 5 and 6 explore the relationship between motion and language production mechanisms, based on four factors: (a) motion types (i.e., toward-, away-, neutral-motion), (b) timing between motion and message formulation (i.e., motion before vs. after message starts formulating), (c) message status (i.e., unconstrained vs. constrained message), and (d) language (i.e., English vs. Japanese). The first section of this chapter summarizes the main results of the experiments reported in Chapters 5 and 6. The second section evaluates the implications of these results in respect to production processes. The third section identifies some limitations of this study, and the last section concludes the dissertation by suggesting the broader benefits of this study.
7.1 Motion, Embodied Cognition, and Message Formulation

This dissertation examined whether, in addition to the speaker’s intended message, other unintended, unplanned, non-linguistic factors, such as motor activities, would affect ongoing cognitive processes, thereby shaping the resulting linguistic output (Chapter 5) or influencing the event description processes (Chapter 6), in English and Japanese. The flexible word order of Japanese was utilized as a mediator to reflect speakers’ active engagement of mental simulation in formulating and producing a language.

In Chapter 5, a series of production experiments aims to answer the following questions: (1) Does the unscripted message construction process (e.g., establishing the relationships among an event’s participants and objects) integrate the speaker’s previous and concurrent physical actions through dynamic motor simulations? and (2) Do physical motions have differing effects on linguistic content depending on when they are executed during the message construction process?

As the set of experiments described in Chapter 5 demonstrated, when speakers are instructed to come up with simple sentences describing possible motor actions using two objects, they need to specify the relational meaning of those objects in order to produce utterances. Results indicate that speakers are influenced by their physical experiences because motor activities are unconsciously captured as an embodied source to establish the relational meaning in the event configuration processes.

The experiments in Chapter 5 are informative and illuminating in regard to the influential role of physical motion when no relationship among the entities in the event has been conceptually established. These experiments alone, however, do not answer the
questions of whether motor information is incorporated into message construction only when messages lack specific internal event structures or whether the sentence processors consistently interact with motor information regardless of their message status. To address such questions, the set of corresponding experiments that are presented in Chapter 6 examined motor effects on constrained message formulation. These experiments expand knowledge of the scope of motor effects on the production process by eliciting constrained messages.

The experiments in Chapter 6 may begin to answer a fundamental question regarding the functional role of motor information. One hypothesis is that this functional role is to be merely beneficial to speakers in generating messages. An alternative hypothesis is that, instead, motor information is unconditionally influential on the speakers and the messages they generate. If the first hypothesis is correct, and the primary functional role of motor information is to aid or assist speakers to generate a message, then speakers utilize the most available motor knowledge in order to fill out the missing relational meaning in the message. In other words, no motor effects should be observed on the constrained message (with its specific relational meaning) because motor information is no longer useful in creating such a message. On the other hand, if the second hypothesis is correct, and speakers are essentially perceptive to motor information that is strongly activated in their cognitive status, then motor information should affect the production of utterances, whether or not such information is beneficial to creating a message.
Results in Chapter 6 showed measurable motor effects on speech onset times to initiate fixed event descriptions, which support the second hypothesis, that is, motor information seems to influence the production processes regardless of message type.

7.2 Implications

Taken together, the findings reported in Chapters 5 and 6 suggest five important implications for our understanding of language production mechanisms in relation to non-linguistic motor experiences. The first implication is that regardless of speakers’ intentions, physical motions influence how speakers construct or frame an event representation of an under-determined message. I assume that after a motor action, the action-path is activated and it remains accessible when speakers seek a particular relational meaning to generate a message whose internal relationship is unconstrained. Conceptually salient and accessible information activated by physical movements is integrated into the message construction process and it shapes the potential utterance. That is, directional movements have an impact on the message planning process, driving people to produce sentences with a particular directional orientation. In earlier work on the production system, the message representation was described as a place where a non-linguistic abstract code is represented in some form. The message was thought to capture the speaker’s intended ideas and meanings and provide the raw material for the next process, grammatical encoding (Bock, 1995). In other words, the message was defined as an intentional representation that influences the process of sentence production. However, this study shows that particular embodied concepts (that are unintentionally activated) can mentally affect speakers in such a way as to immediately influence the process of constructing speech.
The second implication is that speakers are fundamentally sensitive to motor action not only when such motion is critical or necessary to message generation (Experiments 1–4), but also when the relational meaning is fulfilled in the fully determined message (Experiments 5–8). This interactive relationship between motion and utterances with established event structures suggests that speakers adopt motor information or are influenced by motor information not because the language processors are required or forced to specify the relational meaning in generating a message, but because they are by nature responsive to embodied information in the current cognitive status, and thus, constantly interact with salient non-linguistic motor information.

The third implication is that different timing of motor activity relative to scene apprehension followed by meaning construction influences production effects in different ways. First, for messages with no relational meaning, the incorporation of motor activity into message construction has a stronger influence on the message when motion becomes cognitively activated while the message is being formulated (Experiments 3–4) than when motion has been activated prior to the message formulation (Experiments 1–2). This may reflect a characteristic of motor information persistency, that is, motor information that activates general directionality knowledge seems to decay, with the result that its impact on the modulation of event structures decreases over time.

Second, for messages with specific relational meaning (Experiments 5–8), the timing of motor action relative to message formulation changes the components in the production processes that motor information interacts with, and this ultimately produces different motor effects on speech initiation times (i.e., inhibition or facilitation). For example, when physical movements activate general directional information first before
linguistic encoding of the perceived event (Experiments 5–6), then *conceptual specificity* of motor information and event-specific information influences the speech onset times. More specifically, subtle differences in the details of motor information between two pieces of conceptual information (i.e., general directional information activated by hand motions vs. event-specific directional information activated by picture perception) cause inhibition effects due to their non-integratability. In this case, motor information is interacting with conceptual components because it remains cognitively active when a depicted event starts unfolding its information and speakers attempt to extract the meaning.

In contrast, when the event has been completely established first, and motor information is activated second (Experiments 7–8), then consistency of *attention order (motion)* with *word order* seems to be the factor that determines speech initiation times. That is to say, across languages and regardless of the word order (i.e., DO-Loc in English and Loc-DO in Japanese canonical word order), facilitations of speech initiation times are generally observed when the attention order generated by specific directional motion and the word order follow the same accessing order (i.e., a toward-event with Loc-DO word order, and an away-event with DO-Loc word order). The reverse is true for inhibition effects. This indicates that motion is no longer interacting or communicating with conceptual elements, but interacting with the linguistic segments in the subsequent word order/grammatical encoding. This is because the conceptual processes such as meaning extraction should have already been completed by the time speakers perform the physical action. Since the message has been completed, motion interacts with the next process, that is, grammatical encoding that assigns the meanings to the appropriate word order.
In short, physical activities seem to be consumed as a resource for advancing the subsequent linguistic encoding processes.

The fourth implication of motor effects on utterances can be observed from word order in Japanese. The fairly flexible word order reflects how attention order knowledge (i.e., attention order is object-location for away-events and location-object for toward-events) activated by toward or away physical motion modulates the message and/or aligns the message into a particular order in the sentence constructions. As the Isomorphic Mapping Hypothesis (IMS) (O’Grady & Lee, 2005) predicts, throughout the experiments, toward- or away-events are expressed in a temporally consistent word order, that is, canonical or scrambled word order, respectively. If these tight relationships between event language and word order are simply an off-line correlation, then the proportion of event types that speakers select and the correlation between event language and word order should remain the same regardless of the motion speakers performed. Crucially, however, such event and word order consistency becomes more significant after a corresponding motion than after neutral motion.

This indicates that events are more likely expressed in a particular word order, not only because the language processor arranges the word order by tracing/reflecting how the event unfolds, but also because physical motion that activates general embodied knowledge and evokes sensorimotor simulation additionally contributes to word order determination.

The final implication of the experiments in this study is in regard to the degree of motor effects on the sentence production process. It is important to note that the intensity or degree of motion effects on message formulation is different depending on the
presence or absence of relational meaning in the message. For example, when the
relational meaning is under-determined, the motion effectively fills in the missing
information in order to create an event structure. Therefore, the motor effects on the
unconstrained messages are reliably consistent and significant. However, when such
information is already present in the message, the effects are more likely to be small, as
well as inconsistent across language types (i.e., toward- vs. away-language) and
languages (i.e., English and Japanese). These results surely indicate the interactive
relationship between motion and language, and subsequent future work may more
precisely illuminate how motor information interacts with fixed messages in the
production systems.

7.3 Limitations

This study attempts to delineate how a non-linguistic factor, namely body motion,
can fit into production mechanisms and interact with speakers’ thoughts and mental
simulation. The results discussed here contribute to knowledge of the potential nature of
production systems that interact with motor actions in various ways. Nevertheless, a
number of questions still remain unanswered, suggesting promising avenues for future
research. First, because the current research specifically examines the effects of motor
activities through relational meaning within the domain of sentences involving toward- or
away-motions, much work remains to be done with a greater variety of utterances,
including investigating the role of physical motion in producing spontaneous speech.
Future study should also consider utterances that involve other kinds of motor actions in
addition to toward- and away-motion. Toward- and away-motion are not the only
relational meaning possible in sentences, but rather, are one of the options for
determining the relationship among entities. That is, the relational meaning of an event
(i.e., “who did what to whom”) can be encoded by other factors. Investigating motor
effects on different types of directional sentences (e.g., up or down spatial language) or
non-motion sentences would help elucidate the interactive nature of the production
systems more precisely.

Second, throughout the experiments, motion seems to have had considerably
stronger effects on away-language than on toward-language. This may imply that away-
motion is cognitively more distinctive than other motions (i.e., toward-, right-, or left-
motions). However, if this is so, further research is needed to identify the precise reasons
that away-language is more vulnerable than toward-language to the effects of non-
linguistic motor actions.

Third, although the experiments provide compelling cross-linguistic evidence in
support of motion integration processes in constructing messages, such motor effects
become incoherent and weak across languages in some cases. These variable cross-
linguistic findings may be able to be accounted for by the structural differences (i.e., head
initial vs. head final language) or the differences in word order flexibility (i.e., fairly rigid
vs. flexible word order) between the two languages. Motor information may interact with
production processes differently depending on specific features of the language when the
event structure of the message is determined, but this needs further exploration.
7.4 Conclusion

Previous interdisciplinary collaboration between linguistics and cognitive sciences has implicated the use of perceptual and motor knowledge in simulation to facilitate language comprehension. Additionally, physical engagement has been identified as an influential factor in comprehension. The present study attempts to fill a gap between the studies of sentence production and the studies of mental simulation by employing non-linguistic manipulation of bodily motion in order to investigate motion’s effects on embodied cognition and speech construction. Because previous language simulation research deals only with comprehension or evidence from gestures, findings concerning the role of simulation in language production, with data of language content choice, word order, and speech onset time, significantly inform our understanding of simulation processes and meaning encoding.

The central empirical discovery of this research is that, in two typologically distinct languages, both intentionally and unintentionally activated embodied components influence the conceptual or grammatical encodings of the message, resulting in shaping the meaning or influencing the initiation times of the speech output. In other words, messages are determined not just by speakers’ intentions, but also by their current cognitive states. Furthermore, this interaction between motion and established messages indicates that simulation/embodied cognitive information is not a by-product of the sentence production (i.e., simulation would be a by-product if it is performed because of its supplemental role of assisting utterances), but plays a necessary role in the production mechanisms. This suggests that our language is cognitively grounded and acquired through a tight bond with embodied experiences. Moreover, the results from Japanese
may implicate word order as a novel and clear indicator of the existence of simulation, a result that cannot be readily obtained in English.

It must be taken into consideration that these results occurred in a limited way, under controlled laboratory conditions where participants were led to produce utterances without any communicative purpose. The utterances that were produced in this study are different from natural, spontaneous speech. Nevertheless, by showing how embodied knowledge interacts with speech formulation within the existing production model, these results shed light on the important role of embodied cognition in human language production. Furthermore, they advocate for the value of this line of investigation in future research in order to enhance our understanding of how our cognitive systems are grounded in the rich non-linguistic environment in which linguistic acts take place.
APPENDIX A

EXPERIMENTS 1-4: CRITICAL PICTURES

<table>
<thead>
<tr>
<th></th>
<th><img src=".." alt="Image" /></th>
<th><img src=".." alt="Image" /></th>
<th><img src=".." alt="Image" /></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><img src=".." alt="Image" /></td>
<td><img src=".." alt="Image" /></td>
<td><img src=".." alt="Image" /></td>
</tr>
<tr>
<td>2</td>
<td><img src=".." alt="Image" /></td>
<td><img src=".." alt="Image" /></td>
<td><img src=".." alt="Image" /></td>
</tr>
<tr>
<td>3</td>
<td><img src=".." alt="Image" /></td>
<td><img src=".." alt="Image" /></td>
<td><img src=".." alt="Image" /></td>
</tr>
<tr>
<td>4</td>
<td><img src=".." alt="Image" /></td>
<td><img src=".." alt="Image" /></td>
<td><img src=".." alt="Image" /></td>
</tr>
<tr>
<td>5</td>
<td><img src=".." alt="Image" /></td>
<td><img src=".." alt="Image" /></td>
<td><img src=".." alt="Image" /></td>
</tr>
<tr>
<td>6</td>
<td><img src=".." alt="Image" /></td>
<td><img src=".." alt="Image" /></td>
<td><img src=".." alt="Image" /></td>
</tr>
<tr>
<td>7</td>
<td><img src=".." alt="Image" /></td>
<td><img src=".." alt="Image" /></td>
<td><img src=".." alt="Image" /></td>
</tr>
<tr>
<td>8</td>
<td><img src=".." alt="Image" /></td>
<td><img src=".." alt="Image" /></td>
<td><img src=".." alt="Image" /></td>
</tr>
<tr>
<td>9</td>
<td><img src=".." alt="Image" /></td>
<td><img src=".." alt="Image" /></td>
<td><img src=".." alt="Image" /></td>
</tr>
<tr>
<td>10</td>
<td><img src=".." alt="Image" /></td>
<td><img src=".." alt="Image" /></td>
<td><img src=".." alt="Image" /></td>
</tr>
</tbody>
</table>
### APPENDIX B

**EXPERIMENT 2 & 4 (JAPANESE): FILLER PICTURES**

<table>
<thead>
<tr>
<th></th>
<th><img src="image1.png" alt="Image 1" /></th>
<th><img src="image2.png" alt="Image 2" /></th>
<th><img src="image3.png" alt="Image 3" /></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><img src="image4.png" alt="Image 4" /></td>
<td><img src="image5.png" alt="Image 5" /></td>
<td><img src="image6.png" alt="Image 6" /></td>
</tr>
<tr>
<td>2</td>
<td><img src="image7.png" alt="Image 7" /></td>
<td><img src="image8.png" alt="Image 8" /></td>
<td><img src="image9.png" alt="Image 9" /></td>
</tr>
<tr>
<td>3</td>
<td><img src="image10.png" alt="Image 10" /></td>
<td><img src="image11.png" alt="Image 11" /></td>
<td><img src="image12.png" alt="Image 12" /></td>
</tr>
<tr>
<td>4</td>
<td><img src="image13.png" alt="Image 13" /></td>
<td><img src="image14.png" alt="Image 14" /></td>
<td><img src="image15.png" alt="Image 15" /></td>
</tr>
<tr>
<td>5</td>
<td><img src="image16.png" alt="Image 16" /></td>
<td><img src="image17.png" alt="Image 17" /></td>
<td><img src="image18.png" alt="Image 18" /></td>
</tr>
<tr>
<td>6</td>
<td><img src="image19.png" alt="Image 19" /></td>
<td><img src="image20.png" alt="Image 20" /></td>
<td><img src="image21.png" alt="Image 21" /></td>
</tr>
<tr>
<td>7</td>
<td><img src="image22.png" alt="Image 22" /></td>
<td><img src="image23.png" alt="Image 23" /></td>
<td><img src="image24.png" alt="Image 24" /></td>
</tr>
<tr>
<td>8</td>
<td><img src="image25.png" alt="Image 25" /></td>
<td><img src="image26.png" alt="Image 26" /></td>
<td><img src="image27.png" alt="Image 27" /></td>
</tr>
</tbody>
</table>
## APPENDIX C

**EXPERIMENT 1 & 3 (ENGLISH): FILLER PICTURES**

<table>
<thead>
<tr>
<th></th>
<th><img src="image1.png" alt="Image" /></th>
<th><img src="image2.png" alt="Image" /></th>
<th><img src="image3.png" alt="Image" /></th>
<th><img src="image4.png" alt="Image" /></th>
<th><img src="image5.png" alt="Image" /></th>
<th><img src="image6.png" alt="Image" /></th>
<th><img src="image7.png" alt="Image" /></th>
<th><img src="image8.png" alt="Image" /></th>
</tr>
</thead>
<tbody>
<tr>
<td>19</td>
<td>Tomato</td>
<td>20</td>
<td>Knife</td>
<td>Pencil</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----</td>
<td>--------</td>
<td>----</td>
<td>-------</td>
<td>--------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>Blue paper</td>
<td>22</td>
<td>Scissors</td>
<td>Notebooks</td>
<td>Stapler</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>Door</td>
<td>24</td>
<td>Key</td>
<td>Crib</td>
<td>Baby</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX D

EXPERIMENT 1 & 3 (ENGLISH): NORMING FOR PICTURES

Please complete a sentence by describing actions with the two pictured objects. Indicate the direction of the action with an arrow.

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><img src="image1.png" alt="image" /></td>
<td><img src="image2.png" alt="image" /></td>
<td>I am</td>
</tr>
<tr>
<td>2</td>
<td><img src="image4.png" alt="image" /></td>
<td><img src="image5.png" alt="image" /></td>
<td>I am</td>
</tr>
<tr>
<td>3</td>
<td><img src="image7.png" alt="image" /></td>
<td><img src="image8.png" alt="image" /></td>
<td>I am</td>
</tr>
<tr>
<td>4</td>
<td><img src="image10.png" alt="image" /></td>
<td><img src="image11.png" alt="image" /></td>
<td>I am</td>
</tr>
<tr>
<td>5</td>
<td><img src="image13.png" alt="image" /></td>
<td><img src="image14.png" alt="image" /></td>
<td>I am</td>
</tr>
<tr>
<td></td>
<td><img src="image1.png" alt="Image" /></td>
<td><img src="image2.png" alt="Image" /></td>
<td>I am</td>
</tr>
<tr>
<td>---</td>
<td>---------------------</td>
<td>---------------------</td>
<td>--------</td>
</tr>
<tr>
<td>6</td>
<td><img src="image3.png" alt="Image" /></td>
<td><img src="image4.png" alt="Image" /></td>
<td>I am</td>
</tr>
<tr>
<td>7</td>
<td><img src="image5.png" alt="Image" /></td>
<td><img src="image6.png" alt="Image" /></td>
<td>I am</td>
</tr>
<tr>
<td>8</td>
<td><img src="image7.png" alt="Image" /></td>
<td><img src="image8.png" alt="Image" /></td>
<td>I am</td>
</tr>
<tr>
<td>9</td>
<td><img src="image9.png" alt="Image" /></td>
<td><img src="image10.png" alt="Image" /></td>
<td>I am</td>
</tr>
<tr>
<td>10</td>
<td><img src="image11.png" alt="Image" /></td>
<td><img src="image12.png" alt="Image" /></td>
<td>I am</td>
</tr>
<tr>
<td></td>
<td>Image</td>
<td>Image</td>
<td>I am</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>11</td>
<td><img src="image1.png" alt="Image" /></td>
<td><img src="image2.png" alt="Image" /></td>
<td>I am</td>
</tr>
<tr>
<td>12</td>
<td><img src="image3.png" alt="Image" /></td>
<td><img src="image4.png" alt="Image" /></td>
<td>I am</td>
</tr>
<tr>
<td>13</td>
<td><img src="image5.png" alt="Image" /></td>
<td><img src="image6.png" alt="Image" /></td>
<td>I am</td>
</tr>
<tr>
<td>14</td>
<td><img src="image7.png" alt="Image" /></td>
<td><img src="image8.png" alt="Image" /></td>
<td>I am</td>
</tr>
<tr>
<td>15</td>
<td><img src="image9.png" alt="Image" /></td>
<td><img src="image10.png" alt="Image" /></td>
<td>I am</td>
</tr>
<tr>
<td></td>
<td><img src="182x640" alt="Image" /></td>
<td><img src="201x720" alt="Image" /></td>
<td>I am</td>
</tr>
<tr>
<td>---</td>
<td>----------------</td>
<td>-----------------</td>
<td>------</td>
</tr>
<tr>
<td>16</td>
<td><img src="248x562" alt="Image" /></td>
<td><img src="360x720" alt="Image" /></td>
<td>I am</td>
</tr>
<tr>
<td>17</td>
<td><img src="155x559" alt="Image" /></td>
<td><img src="213x615" alt="Image" /></td>
<td>I am</td>
</tr>
<tr>
<td>18</td>
<td><img src="275x481" alt="Image" /></td>
<td><img src="342x548" alt="Image" /></td>
<td>I am</td>
</tr>
<tr>
<td>19</td>
<td><img src="152x472" alt="Image" /></td>
<td><img src="228x548" alt="Image" /></td>
<td>I am</td>
</tr>
<tr>
<td>20</td>
<td><img src="269x222" alt="Image" /></td>
<td><img src="350x321" alt="Image" /></td>
<td>I am</td>
</tr>
<tr>
<td>21</td>
<td><img src="155x136" alt="Image" /></td>
<td><img src="213x194" alt="Image" /></td>
<td>I am</td>
</tr>
<tr>
<td>22</td>
<td><img src="image1.png" alt="Image" /></td>
<td><img src="image2.png" alt="Image" /></td>
<td>I am</td>
</tr>
<tr>
<td>23</td>
<td><img src="image3.png" alt="Image" /></td>
<td><img src="image4.png" alt="Image" /></td>
<td>I am</td>
</tr>
<tr>
<td>24</td>
<td><img src="image5.png" alt="Image" /></td>
<td><img src="image6.png" alt="Image" /></td>
<td>I am</td>
</tr>
<tr>
<td>25</td>
<td><img src="image7.png" alt="Image" /></td>
<td><img src="image8.png" alt="Image" /></td>
<td>I am</td>
</tr>
<tr>
<td>26</td>
<td><img src="image9.png" alt="Image" /></td>
<td><img src="image10.png" alt="Image" /></td>
<td>I am</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>27</td>
<td><img src="image1.png" alt="Image" /></td>
<td>I am</td>
<td></td>
</tr>
<tr>
<td>28</td>
<td><img src="image2.png" alt="Image" /></td>
<td>I am</td>
<td></td>
</tr>
<tr>
<td>29</td>
<td><img src="image3.png" alt="Image" /></td>
<td>I am</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td><img src="image4.png" alt="Image" /></td>
<td>I am</td>
<td></td>
</tr>
<tr>
<td>31</td>
<td><img src="image5.png" alt="Image" /></td>
<td>I am</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td><img src="image1.png" alt="Log" /> <img src="image2.png" alt="Axe" /></td>
<td>I am</td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>--------------------------------------</td>
<td>------</td>
<td></td>
</tr>
<tr>
<td>33</td>
<td><img src="image3.png" alt="Fruit" /> <img src="image4.png" alt="Blender" /></td>
<td>I am</td>
<td></td>
</tr>
<tr>
<td>34</td>
<td><img src="image5.png" alt="Toast" /> <img src="image6.png" alt="Knife" /></td>
<td>I am</td>
<td></td>
</tr>
<tr>
<td>35</td>
<td><img src="image7.png" alt="Boat Shoe" /> <img src="image8.png" alt="Shoe Horn" /></td>
<td>I am</td>
<td></td>
</tr>
<tr>
<td>36</td>
<td><img src="image9.png" alt="Wine Bottle" /> <img src="image10.png" alt="Wine Glass" /></td>
<td>I am</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>37</td>
<td></td>
<td></td>
<td>I am</td>
</tr>
<tr>
<td>38</td>
<td></td>
<td></td>
<td>I am</td>
</tr>
<tr>
<td>39</td>
<td></td>
<td></td>
<td>I am</td>
</tr>
<tr>
<td>40</td>
<td></td>
<td></td>
<td>I am</td>
</tr>
<tr>
<td>41</td>
<td></td>
<td></td>
<td>I am</td>
</tr>
<tr>
<td>42</td>
<td></td>
<td></td>
<td>I am</td>
</tr>
<tr>
<td>43</td>
<td><img src="image" alt="Notepad" /></td>
<td><img src="image" alt="Stapler" /></td>
<td>I am</td>
</tr>
<tr>
<td>44</td>
<td><img src="image" alt="Cake" /></td>
<td><img src="image" alt="Fork" /></td>
<td>I am</td>
</tr>
<tr>
<td>45</td>
<td><img src="image" alt="Flowers" /></td>
<td><img src="image" alt="Watering can" /></td>
<td>I am</td>
</tr>
<tr>
<td>46</td>
<td><img src="image" alt="Door" /></td>
<td><img src="image" alt="Key" /></td>
<td>I am</td>
</tr>
<tr>
<td>47</td>
<td><img src="image" alt="Paper clips" /></td>
<td></td>
<td>I am</td>
</tr>
<tr>
<td>48</td>
<td>![Image of a pencil and a sharpener]</td>
<td>I am</td>
<td></td>
</tr>
</tbody>
</table>
2つの写真を使って、何かの出来事を表す文章を完成させて下さい。
そしてその動作が自分から見てどの方向なのかを矢印で示して下さい。

<table>
<thead>
<tr>
<th>1</th>
<th>左</th>
<th>私は</th>
<th>右</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>右</td>
<td>私は</td>
<td>左</td>
</tr>
<tr>
<td>3</td>
<td>左</td>
<td>私は</td>
<td>右</td>
</tr>
<tr>
<td>4</td>
<td>前</td>
<td>私は</td>
<td>後</td>
</tr>
<tr>
<td>5</td>
<td>右</td>
<td>私は</td>
<td>左</td>
</tr>
<tr>
<td>6</td>
<td>右</td>
<td>私は</td>
<td>左</td>
</tr>
<tr>
<td>7</td>
<td>苹果</td>
<td>私は</td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>冰箱 牛奶</td>
<td>私は</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>手表</td>
<td>私は</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>毛巾 纸巾</td>
<td>私は</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>首饰 珍珠戒指</td>
<td>私は</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>脚 袜子</td>
<td>私は</td>
<td></td>
</tr>
<tr>
<td>ページ</td>
<td>イラスト</td>
<td>言葉</td>
<td>ビジュアル</td>
</tr>
<tr>
<td>--------</td>
<td>----------</td>
<td>------</td>
<td>-----------</td>
</tr>
<tr>
<td>13</td>
<td>帽子</td>
<td>私は</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>赤ワイン</td>
<td>私は</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>洗濯機</td>
<td>私は</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>鉛筆立てペン</td>
<td>私は</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>鍵</td>
<td>私は</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>17</td>
<td>18</td>
<td>20</td>
</tr>
<tr>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
</tr>
<tr>
<td><img src="image1.png" alt="書架" /></td>
<td><img src="image2.png" alt="書籍" /></td>
<td><img src="image3.png" alt="私は" /></td>
<td><img src="image4.png" alt="私は" /></td>
</tr>
<tr>
<td><img src="image6.png" alt="クリスマスツリー" /></td>
<td><img src="image7.png" alt="ボタン" /></td>
<td><img src="image8.png" alt="私は" /></td>
<td><img src="image9.png" alt="私は" /></td>
</tr>
<tr>
<td><img src="image11.png" alt="食器棚" /></td>
<td><img src="image12.png" alt="ナイフ" /></td>
<td><img src="image13.png" alt="私は" /></td>
<td><img src="image14.png" alt="私は" /></td>
</tr>
<tr>
<td>22</td>
<td>奥のコンセント</td>
<td>私は</td>
<td>しまはありません</td>
</tr>
<tr>
<td>26</td>
<td>女性の顔と手首のアクセサリー</td>
<td>私は</td>
<td>しまはありません</td>
</tr>
<tr>
<td>23</td>
<td>子供のベビー服</td>
<td>私は</td>
<td>しまはありません</td>
</tr>
<tr>
<td>24</td>
<td>洗濯機とリネンタオル</td>
<td>私は</td>
<td>しまはありません</td>
</tr>
<tr>
<td>25</td>
<td>バスケットボールのラックとバスケット</td>
<td>私は</td>
<td>しまはありません</td>
</tr>
<tr>
<td>27</td>
<td>28</td>
<td>29</td>
<td>30</td>
</tr>
<tr>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td><img src="image1" alt="" /></td>
<td><img src="image2" alt="" /></td>
<td><img src="image3" alt="" /></td>
<td><img src="image4" alt="" /></td>
</tr>
<tr>
<td>は</td>
<td>は</td>
<td>は</td>
<td>は</td>
</tr>
<tr>
<td>32</td>
<td></td>
<td>私は</td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX F

CRITICAL PICTURES IN EXPERIMENTS 5-8

<table>
<thead>
<tr>
<th></th>
<th>Toward-motion events</th>
<th>Away-motion events</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><img src="image1.png" alt="Image" /></td>
<td><img src="image2.png" alt="Image" /></td>
</tr>
<tr>
<td>2</td>
<td><img src="image3.png" alt="Image" /></td>
<td><img src="image4.png" alt="Image" /></td>
</tr>
<tr>
<td>3</td>
<td><img src="image5.png" alt="Image" /></td>
<td><img src="image6.png" alt="Image" /></td>
</tr>
<tr>
<td>4</td>
<td><img src="image7.png" alt="Image" /></td>
<td><img src="image8.png" alt="Image" /></td>
</tr>
</tbody>
</table>
**APPENDIX G**

**FILLER PICTURES IN EXPERIMENTS 5-8**

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><img src="image1" alt="Image" /></td>
<td><img src="image2" alt="Image" /></td>
<td><img src="image3" alt="Image" /></td>
</tr>
<tr>
<td>2</td>
<td><img src="image4" alt="Image" /></td>
<td><img src="image5" alt="Image" /></td>
<td><img src="image6" alt="Image" /></td>
</tr>
<tr>
<td>3</td>
<td><img src="image7" alt="Image" /></td>
<td><img src="image8" alt="Image" /></td>
<td><img src="image9" alt="Image" /></td>
</tr>
<tr>
<td>4</td>
<td><img src="image10" alt="Image" /></td>
<td><img src="image11" alt="Image" /></td>
<td><img src="image12" alt="Image" /></td>
</tr>
<tr>
<td>5</td>
<td><img src="image13" alt="Image" /></td>
<td><img src="image14" alt="Image" /></td>
<td><img src="image15" alt="Image" /></td>
</tr>
<tr>
<td>6</td>
<td><img src="image16" alt="Image" /></td>
<td><img src="image17" alt="Image" /></td>
<td><img src="image18" alt="Image" /></td>
</tr>
<tr>
<td>7</td>
<td><img src="image19" alt="Image" /></td>
<td><img src="image20" alt="Image" /></td>
<td><img src="image21" alt="Image" /></td>
</tr>
<tr>
<td>8</td>
<td><img src="image22" alt="Image" /></td>
<td><img src="image23" alt="Image" /></td>
<td><img src="image24" alt="Image" /></td>
</tr>
<tr>
<td>9</td>
<td><img src="image25" alt="Image" /></td>
<td><img src="image26" alt="Image" /></td>
<td><img src="image27" alt="Image" /></td>
</tr>
<tr>
<td>10</td>
<td><img src="image28" alt="Image" /></td>
<td><img src="image29" alt="Image" /></td>
<td><img src="image30" alt="Image" /></td>
</tr>
</tbody>
</table>
REFERENCES


Electrophysiological evidence for the primacy of shape over color in object

Pulvermüller, F., Haerle, M., & Hummel, F. (2001). Walking or talking?: Behavioral and
neurophysiological correlates of action verb processing. *Brain and Language, 78*,
143–168.


The role of lexical movements in speech production. *Psychological Science, 7*,
226–231.

Richardson, D. C., Spivey, M. J., Barsalou, L. W., & McRae, K. (2003). Spatial
representations activated during real-time comprehension of verbs. *Cognitive

Cognition, 70*, 253–258.

Gumperz & S. Levinson (Eds.), *Rethinking linguistic relativity* (pp. 70–96).
Cambridge: Cambridge University Press.


477–488.


