Interpretation of Overt Pronouns in L1 and L2 Japanese: The Role of Context

Marisa A. Nagano

Graduate Center, City University of New York

How does access to this work benefit you? Let us know!

Follow this and additional works at: https://academicworks.cuny.edu/gc_etds

Part of the Linguistics Commons

Recommended Citation


This Dissertation is brought to you by CUNY Academic Works. It has been accepted for inclusion in All Dissertations, Theses, and Capstone Projects by an authorized administrator of CUNY Academic Works. For more information, please contact deposit@gc.cuny.edu.
INTERPRETATION OF OVERT PRONOUNS IN L1 AND L2 JAPANESE: THE ROLE OF CONTEXT

by

MARISA NAGANO

A dissertation submitted to the Graduate Faculty in Linguistics in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York.

2015
This manuscript has been read and accepted by the Graduate Faculty in Linguistics in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

Date
Professor Gita Martohardjono
Chair of Examining Committee

Date
Professor Gita Martohardjono
Executive Officer

Professor Gita Martohardjono
Professor William McClure
Professor Ricardo Otheguy
Professor Richard Schwartz

Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK
ABSTRACT

INTERPRETATION OF OVERT PRONOUNS IN L1 AND L2 JAPANESE: THE ROLE OF CONTEXT

by

Marisa Nagano

Advisor: Dr. Gita Martohardjono

This dissertation examines the interpretation of Japanese overt 3rd person pronouns by native (N=21) and advanced L1-English/L2-Japanese (N=20) speakers, using interpretation, reaction-time, and eye-tracking data. Previous studies have investigated L1 and L2 pronoun interpretation in null-subject languages like Spanish, Italian, and Greek, finding that L1 speakers tend to show a topic-shift effect for overt pronouns more often than L2ers (Sorace, 2011). Similar studies on Japanese, which allows null subjects but also differs greatly from these languages, have been mixed; crucially, several studies found no antecedent bias for overt pronouns by L1 speakers to begin with (Okuma, 2012; Ueno & Kehler, 2010).

This study first argues that Japanese overt 3rd person pronouns differ crucially from those in Italian-like languages, in that they do not elicit a generalized conversational implicature for topic-shift, nor do they fit on an anaphora hierarchy. Instead, Japanese 3rd person pronouns are better analyzed as triggering particularized conversational implicatures, i.e., that their interpretation must be calculated on an utterance-by-utterance basis based on the specific context of each utterance (as opposed to generalized conversational implicatures, which can be cancelled by context but are not created by it). This study further argues that advanced L2 speakers...
therefore face a different task than L2 speakers of Italian-like languages, and crucially must inhibit English in order to achieve a native-like interpretation pattern.

These arguments are backed up by experimental data, in which the addition of a context specifically supporting topic-shift succeeds in eliciting a consistent topic-shift interpretation for 3rd person overt pronouns in native Japanese speakers. Processing data, including eye-tracking data, support the idea that the interpretation of overt 3rd person pronouns is calculated from context, with little processing penalty regardless of ultimate interpretation (in contrast with a strong processing cost for choosing the dispreferred interpretation for nulls).

L2 participants in this study showed interpretation patterns qualitatively different from those of native speakers, as opposed to L2 Italian speakers in the studies cited above, who showed quantitatively different but qualitatively similar patterns to native speakers. Individual L2 interpretation data in this study, however, revealed bimodality: a minority of L2 participants patterned with the L1 mean, whereas the rest patterned much lower. Eye-tracking data further suggest that L2 participants begin to process the overt pronoun in a way similar to L1 participants, but quickly fall back onto an English-like interpretation. Together, this data suggests that if L2 participants are successful in inhibiting an English-like response to the pronoun, they will then switch over to a native-like interpretation pattern (without the quantitative differences found in L2 speakers of other languages). This effect is not unexpected under the particularized conversational implicature approach to Japanese overt 3rd person pronouns, in which interpretation falls out directly from the Cooperative Principle (Grice, 1975), rather than a particular setting on an anaphora hierarchy.
ACKNOWLEDGEMENTS

The dissertation was written with the support and advice of countless others. My committee, Dr. Gita Martohardjono, Dr. William McClure, Dr. Ricardo Otheguy, and Dr. Richard Schwartz, was of course a source of invaluable guidance and assistance. This project began as a term paper in Dr. Martohardjono’s Second Language Acquisition class, and she has guided me through many years (including both my second qualifying paper and this dissertation) as I developed it into a full-fledged experimental study. I am very grateful for her advice on both content and logistics, and especially for encouraging me to use eye-tracking, a new methodology for me that ended up providing a great deal of fascinating data and ideas for future studies.

I am very grateful to Dr. McClure for providing me with helpful insights into semantics and Japanese linguistics, and for supporting my research ever since my first qualifying paper. Dr. Otheguy provided a crucial point of view as a scholar of Romance languages, and helped me to examine the study of overt and null pronouns from several different theoretical perspectives. I am grateful to Dr. Schwartz for both the use of his eye-tracking lab for testing L2 participants, and for helpful insights into my data analysis.

I owe special thanks to Zara Waldman and Karece Lopez from Dr. Schwartz’ lab, for indispensable help with E-Prime, Tobii, and Matlab. Many thanks as well to everyone in Dr. Martohardjono’s Second Language Acquisition lab, for feedback on conference presentations and on earlier versions of this project. Additionally, I am grateful to Luca Campanelli for invaluable help with both R and my statistical analysis in general. Ankit Rastogi also provided much-needed statistics help.

Deep thanks go to Dr. Yuki Hirose for the use of her lab at the University of Tokyo, where I collected my control data, without which I would have no study. Thanks also to Dr.
Doug Roland for crucial E-Prime debugging help while I was there, and to Dr. Baris Kahraman for pointing me towards helpful literature on Japanese pronouns. Thanks also to Chie Nakamura for running two L2 participants for me in the Tokyo lab after I left.

I am extremely grateful to all of my fellow students in the Linguistics Department at the Graduate Center. The collegial, friendly atmosphere among students in the Department was a major source of support during the long path to degree. Thanks especially to Emily Zane, Teresa O’Neill, Edmund O’Neill, Erin Quirk, Sonya Elliot, Katie Hawkland, Sarah Kresh, and Elizabeth Pratt. I wouldn’t have made it without you.

I am forever indebted to my parents, Maryann and Larry Genuardi, who have unwaveringly supported my education for as long as I can remember (sorry it took so long!). Thanks also to my sisters, Mia and Monica Genuardi, for their constant support. I am also grateful to both my parents and my sisters for two-plus years of babysitting services, and especially to Mia for being the best nanny in the world.

I am grateful to my late mother-in-law, Sachiyó Nagano, for boosting my confidence in my Japanese abilities, and for giving me a home in Japan. And of course, this project would not have been completed without the help of my husband, Dr. Tomonori Nagano, who listened to all of my worries and concerns and offered his unfailing support, both emotionally and practically (including his services as a native Japanese speaker). Finally, thanks to our son, Kiyo Xavier, who inspired me to finish and whose smile kept me going throughout this whole project.
TABLE OF CONTENTS

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

Chapter 2: Second Language Acquisition of Overt Pronouns ..................................................4

  2.1 Background: The Interface Hypothesis ................................................................. 4

  2.2 Previous Interface Hypothesis Studies .................................................................. 5

    2.2.1 Childhood Bilingualism .................................................................................. 5

    2.2.2 Adult Bilingual Populations ......................................................................... 6

    2.2.3 Explaining Monolingual and Bilingual Differences ....................................... 9

    2.2.4 Summary ........................................................................................................ 11

  2.3 Overt and Null Pronouns in L2 Japanese ............................................................... 12

  2.4 This Study ............................................................................................................. 18

Chapter 3: Overt and Null Pronouns in Theory ................................................................ 19

  3.1 Introduction ........................................................................................................... 19

  3.2 Behavior of Overt Pronouns in Null Argument Languages .................................... 19

    3.2.1 Evidence from Psycholinguistic Studies ....................................................... 19

    3.2.2 Evidence from Variationist Studies ................................................................ 21

  3.3 Interpreting Overt and Null Pronouns .................................................................. 23

    3.3.1 Strong and Weak Pronouns ......................................................................... 23

    3.3.2 Anaphora Hierarchies ................................................................................... 24

  3.4 Anaphora Hierarchies and Linguistic Variation ..................................................... 30

    3.4.1 Variation in Anaphor Rankings ..................................................................... 30

    3.4.2 Variation in Antecedent Saliency .................................................................. 32
Chapter 1: Neo-Gricean Theories of Anaphora

3.5 Neo-Gricean Theories of Anaphora

3.5.1 Background: Maxim of Quantity and Anaphora

3.5.2 Particularized and Generalized Conversational Implicatures

3.5.3 A GCI Analysis of Anaphora

3.6 Summary

Chapter 4: Overt Pronouns in Japanese

4.1 Topic-shift effect in Japanese: Does It Exist?

4.2 Previous Studies on Japanese: Little Evidence of Topic-shift

4.2.1 L1 Control Data in SLA Studies

4.2.2 Evidence from Native Speaker Production Studies

4.2.3 Evidence from Corpus Data

4.3 Properties of Overt Pronouns in Japanese

4.3.1 Japanese Pronouns as Nouns

4.3.2 Third-person Pronouns and Demonstratives

4.3.3 Sociopragmatic Constraints on Third-person Pronouns

4.3.4 Infrequency of Overt Pronouns

4.3.5 Summary: 3P Pronouns as Infrequent, Indeterminate Forms

4.4 Capturing the Difference between the Two Language Types: A Neo-Gricean Approach

4.4.1 A PCI Approach for Japanese

4.4.2 Previous Attempts to put kare and kanojo on Anaphora Scales

4.5 Experiment
# Chapter 5: Experimental Set-up: Research Questions, Methods & Predictions

5.1 Research Questions ........................................ 67
5.2 Methods .................................................. 69
   5.2.1 Overview of Experiment .................................. 69
   5.2.2 Test Items ............................................. 69
   5.2.3 Participants .......................................... 72
   5.2.4 Procedure ............................................ 74
   5.2.5 Data Coding ......................................... 75
5.3 Predictions ............................................... 77
   5.3.1 Pronouns Condition ................................... 77
   5.3.2 Particle Condition ................................... 78
   5.3.3 Group Condition .................................... 79

# Chapter 6: Results ........................................ 81
6.1 Introduction ............................................. 81
6.2 Interpretation Data ....................................... 82
   6.2.1 Overview: Interpretation Means ....................... 82
   6.2.2 Inferential Statistics: Interpretation Data ............. 84
   6.2.3 L2 Individual Results, Proficiency, and Language Background
       Survey Data ............................................ 87
6.3 Reaction Time (RT) Data .................................. 92
   6.3.1 Overview: RT Means .................................. 92
   6.3.2 Inferential Statistics: RT Data ......................... 94
6.4 Eye-tracking (ET) Data: Test Sentence

6.4.1 Overview: Test Sentence Time Course Data

6.4.2 Inferential Statistics: ET Data

6.4.2.1 Pronoun (PR)

6.4.2.2 DO2

6.4.2.3 V2

6.4.3 Summary

6.5 Eye-tracking (ET) Data: Question

6.5.1 Overview: Question Time Course Data

6.5.2 Inferential Statistics: ET Data-Question

6.5.2.1 Question

6.5.2.2 After Question (AQ)

6.5.3 Summary: Question and After Question Periods

6.6 Choosing With and Against Prediction

6.6.1 Reaction Time by WP/AP Trials

6.6.2 Eye-tracking Data by WP/AP Trials: L1 Group

6.6.3 Eye-tracking Data by WP/AP Trials: L2 Group

6.7 Summary of Results

6.7.1 Pronoun

6.7.2 Particle

6.7.3 Group

Chapter 7: Discussion

7.1 L1 Group
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.1.1</td>
<td>L1 Interpretation Data</td>
<td>145</td>
</tr>
<tr>
<td>7.1.2</td>
<td>L1 Processing Data</td>
<td>146</td>
</tr>
<tr>
<td>7.2</td>
<td>L2 Group</td>
<td>154</td>
</tr>
<tr>
<td>7.2.1</td>
<td>L2 Interpretation Data</td>
<td>154</td>
</tr>
<tr>
<td>7.2.2</td>
<td>L2 Processing Data</td>
<td>157</td>
</tr>
<tr>
<td>7.3</td>
<td>Explaining L1/L2 Divergence</td>
<td>163</td>
</tr>
<tr>
<td>7.3.1</td>
<td>Nulls</td>
<td>163</td>
</tr>
<tr>
<td>7.3.2</td>
<td>Overts</td>
<td>165</td>
</tr>
<tr>
<td>7.4</td>
<td>Summary of Experiment</td>
<td>168</td>
</tr>
<tr>
<td>7.5</td>
<td>Overall Summary</td>
<td>170</td>
</tr>
</tbody>
</table>

**References** | 173  |

**List of Abbreviations** | 181  |
LIST OF TABLES AND FIGURES

Table 1. Results of Tsimpli et al., 2004 (Forward Anaphora) ____________ 7
Table 2. Results of Tsimpli et al., 2004 (Backward Anaphora) ____________ 7
Table 3. Results of Sorace & Filiaci, 2006 (Forward Anaphora) ____________ 8
Table 4. Results of Sorace & Filiaci, 2006 (Backward Anaphora) ____________ 8
Table 5. Results of Sorace & Filiaci, 2006 (Backward Anaphora) ____________ 14
Table 6. Results of Sorace & Filiaci, 2006 (Native Italian Speakers) ____________ 20
Table 7. Results of Sorace & Filiaci, 2006 and Okuma (2012), Compared ______ 44
Table 8. First- and Second-person Pronouns in Martin (1975) ____________ 52
Figure 1. Example test item with screenshot of computer display ____________ 70
Table 9. Self-ratings of English Proficiency by L1 Japanese Participants ______ 73
Figures 2a, 2b, and 2c. Mean proportion of topic-shift responses by Pronoun
    and Particle (L1); by Pronoun only (L1) and by Particle only (L1) ______ 82
Figures 3a, 3b, and 3c. Mean proportion of topic-shift responses by Pronoun
    and Particle (L2); by Pronoun only (L2) and by Particle only (L2) ______ 84
Figure 4. Interpretation data means _______________________________ 85
Figure 5. R output for final model for interpretation data ____________ 86
Figures 6 & 7. Proportion of topic-shift for overt pronouns (individual results, L2)
    & Proportion of topic-shift for overt pronouns (individual results, L1) ____ 88
Figures 8 & 9. Correlation of proficiency and proportion of topic-shift response
    (L2) & Individual proficiency scores (L2) ________________________________ 89
Figures 10 & 11. Correlation of age of onset of acquisition of Japanese and
    proportion of topic-shift responses (L2) & Correlation of semesters of
formal study of Japanese and proportion of topic-shift responses (L2)_.90

Table 10. Language Background Survey Results for Native-like L2 Participants_.92

Figures 12a, 12b, & 12c. Reaction time by Pronoun and Particle (L1), by
Pronoun (L1), and by Particle (L1)____________________________93

Figures 13a, 13b, & 13c. Reaction time by Pronoun and Particle (L2), by
Pronoun (L2), and by Particle (L2)____________________________94

Figures 14a & 14b. Reaction time means (outliers replaced) & Reaction time
means (outliers replaced, log transformed)________________________95

Figure 15. R output for final model for reaction time data__________________96

Figure 16. Eye-tracking data for Test Sentence (gaNull, L1)______________98
Figure 17. Eye-tracking data for Test Sentence (gaNull, L2)______________98
Figure 18. Eye-tracking data for Test Sentence (waNull, L1)_____________100
Figure 19. Eye-tracking data for Test Sentence (waNull, L2)_____________100
Figure 20. Eye-tracking data for Test Sentence (gaOvert, L1)_____________101
Figure 21. Eye-tracking data for Test Sentence (gaOvert, L2)_____________101
Figure 22. Eye-tracking data for Test Sentence (waOvert, L1)_____________102
Figure 23. Eye-tracking data for Test Sentence (waOvert, L2)_____________102
Figure 24 & Table 11. Mean target advantage score during pronoun in
graph & table form_______________________________104
Figure 25 & Table 12. Mean target advantage score during DO2 in graph
& table form_______________________________105
Figure 26. R output for final model for target advantage data (DO2)________106
Figure 27 & Table 13. Mean target advantage score during V2 in graph & table form

Figure 28. R output for final model for target advantage data (V2)

Figure 29. Eye-tracking data for Question (gaNull, L1)

Figure 30. Eye-tracking data for Question (gaNull, L2)

Figure 31. Eye-tracking data for Question (gaNull-Against Prediction, L1)

Figure 32. Eye-tracking data for Question (gaNull-With Prediction, L1)

Figure 33. Eye-tracking data for Question (waNull, L1)

Figure 34. Eye-tracking data for Question (waNull, L2)

Figure 35. Eye-tracking data for Question (waNull-With Prediction, L1)

Figure 36. Eye-tracking data for Question (waNull-With Prediction, L2)

Figure 37. Eye-tracking data for Question (waOvert, L1)

Figure 38. Eye-tracking data for Question (waOvert, L2)

Figure 39. Eye-tracking data for Question (gaOvert, L1)

Figure 40. Eye-tracking data for Question (gaOvert, L2)

Figure 41. Eye-tracking data for Question (gaOvert-With Prediction, L1)

Figure 42. Eye-tracking data for Question (waOvert-With Prediction, L1)

Figure 43. Eye-tracking data for Question (gaOvert-With Prediction, L2)

Figure 44. Eye-tracking data for Question (waOvert-With Prediction, L2)

Figure 45. Eye-tracking data for Question (gaOvert-Against Prediction, L1)

Figure 46. Eye-tracking data for Question (waOvert-Against Prediction, L1)

Figure 47. Eye-tracking data for Question (gaOvert-Against Prediction, L2)

Figure 48. Eye-tracking data for Question (waOvert-Against Prediction, L2)
Figure 49 & Table 14. Mean target advantage during Question in graph & table form. 122

Figure 50. R output for final model for target advantage data (Question). 123

Figure 51 & Table 15. Mean target advantage during AQ in graph & table form. 124

Figure 52. R output for final model for target advantage data (After Question). 125

Figure 53. Mean reaction time for trials answered against prediction for Pronoun (L1). 129

Figures 54 & 55. Mean reaction time for trials answered with prediction for Pronoun & against prediction for Pronoun. 130

Figure 56. Eye-tracking data for Test Sentence (waNull-Against Prediction, L1). 131

Figure 57. Eye-tracking data for Test Sentence (waNull-With Prediction, L1). 131

Figure 58. Eye-tracking data for Test Sentence (gaOvert-Against Prediction, L1). 132

Figure 59. Eye-tracking data for Test Sentence (gaOvert-With Prediction, L1). 132

Figure 60. Eye-tracking data for Test Sentence (waOvert-With Prediction, L1). 133

Figure 61. Eye-tracking data for Test Sentence (waOvert-Against Prediction, L1). 133

Figure 62. Eye-tracking data for Test Sentence (gaNull-With Prediction, L1). 134

Figure 63. Eye-tracking data for Test Sentence (gaNull-Against Prediction, L1). 134

Figure 64 & 65. L1 Target advantage scores for trials answered against and with prediction for Pronoun (DO2) & (V2). 135

Figure 66. L1 Target advantage scores for trials answered against and with prediction for Pronoun (Question). 136

Figure 67. L1 Target advantage scores for trials answered against and with
prediction for Pronoun (AQ)………………………………………………..137
Figures 68 & 69. L2 Target advantage scores for trials answered against
and with prediction for Pronoun (DO2) & (V2)……………………………..138
Figure 70. Eye-tracking data for Test Sentence (gaOvert-With Prediction, L2)___139
Figure 71. Eye-tracking data for Test Sentence (gaOvert-Against Prediction, L2)___139
Figure 72. Eye-tracking data for Test Sentence (waOvert With Prediction, L2)___140
Figure 73. Eye-tracking data for Test Sentence (waOvert-Against Prediction, L2)___140
Figures 74 & 75. L2 Target advantage scores for trials answered against and with
prediction for Pronoun (Question) & (AQ)…………………………………141
Figure 76. Eye-tracking data for Question (gaNull-Against Prediction, L1)____148
Figure 77. Eye-tracking data for Question (gaNull- With Prediction, L1)____149
Figure 78. Eye-tracking data for Question (gaOvert-With Prediction, L1)_____150
Figure 79. Eye-tracking data for Question (waOvert- With Prediction, L1)_____150
Figure 80. Eye-tracking data for Question (gaOvert-Against Prediction, L1)_____151
Figure 81. Eye-tracking data for Question (waOvert-Against Prediction, L1)_____151
Table 16. Topic-shift rates found in results of Tsimpli, et al., 2004,
Sorace & Filiaci, 2006, Belletti, et al., 2007, and This Study ___________155
Figure 82. Eye-tracking data for Test Sentence (gaOvert-Against Prediction, L2)___159
Figure 83. Eye-tracking data for Test Sentence (waOvert-Against Prediction, L2)___159
Figure 84. Eye-tracking data for Test Sentence (gaOvert- With Prediction, L2)____161
Figure 85. Eye-tracking data for Question (gaOvert- With Prediction, L2)_______161
Figure 86. Eye-tracking data for Test Sentence (waOvert- With Prediction, L2)____162
Figure 87. Eye-tracking data for Question (waOvert- With Prediction, L2)_______162
CHAPTER 1
INTRODUCTION

This study examines the interpretation of overt 3rd person pronouns in Japanese by both native speakers of Japanese and advanced L2 speakers whose native language is English, building off of previous studies on near-native L2 speakers of Italian (and other bilingual populations) testing the Interface Hypothesis (IH), which claims that structures that involve the interface of two linguistic modules remain a source for divergence for monolingual and bilingual groups. Specifically, the IH studies have provided experimental evidence that while monolingual speakers of Italian-like languages tend to interpret overt pronouns as indicating topic-shift (i.e., referring to a non-subject), various bilingual groups do so to a statistically-significant lesser extent (Sorace & Filiaci, 2006; Tsimpli, Sorace, Heycock & Filiaci, 2004; Argyi & Sorace, 2007). However, various experimental studies on Japanese, including one done in the IH mold, have found that Japanese native speakers tend to interpret the overt pronoun as ambiguous between topic-shift and non-topic-shift interpretations (Okuma, 2012; Kanno, 1997; Ueno & Kehler, 2010).

The fact that Japanese and Italian, both languages that allow null subjects, should differ in overt pronoun interpretation is relatively unsurprising. Japanese, English, and Italian represent three types of language: those like English, which prefers its subjects and direct objects to be overt; those like Italian, which can leave subjects unpronounced while containing semantic pronominal information (person and number features) in verb agreement morphology; and those like Japanese, which require neither overt arguments nor agreement morphology. The differences among these languages have long been studied in the literature on the syntax of null subjects, often variations on the theme described above: English and Italian on one hand require an [EPP]
feature to be checked, but go about it in different ways (English with a subject pronoun, Italian with agreement morphology), whereas Japanese has no such feature to be checked in the first place (Chomsky, 1982; Huang, 1984/1989; Jaeggli & Safir, 1989; Zushi, 2003). Previous literature on overt pronoun interpretation, such as various anaphor hierarchies that rank referring expressions along a scale depending on the saliency of the antecedents that they retrieve, also give leeway for languages to vary (Givon, 1983; Ariel, 1990; Gundel, Hedberg & Zacharski, 1993). Recent psycholinguistic studies, including those in the IH mold, have offered further experimental evidence that different anaphors are sensitive to different types of antecedent saliency, both within and across languages (Kaiser & Trueswell, 2008; Filiaci, Sorace & Carreiras, 2013).

In this paper, I argue that differences in overt pronouns in Italian and Japanese trace back to unique properties of Japanese 3rd person pronouns (including their mixed lexical properties and relative diachronic youth), and can best be captured using a Gricean distinction between particularized and generalized conversational implicature (which can itself be integrated into the various anaphora scales described above) (Grice, 1975; Levinson, 2000; Huang, 1994). I then describe an experiment I conducted in which Japanese native-speaking participants gave a consistent (more than chance) topic-shift interpretation for overt pronouns, in contrast to the previous studies cited above. Advanced L2 participants, on the other hand, do not choose an overwhelming topic-shift interpretation for the overt pronoun, with an even greater disparity in interpretation preference between L1 and L2 participants in this study and in the previous IH studies. I argue that this disparity can also be explained by the differences between overt pronouns in Italian and Japanese, which mean that L2 learners of these two languages face qualitatively different processes in arriving at native-like interpretation patterns. Specifically, L2
speakers of Italian primarily rely on input from the L2, while L2 speakers of Japanese must rely solely on decreased input (and thus decreased activation) of the L1.

In Chapter 2, I describe second language acquisition studies of overt pronoun interpretation, including various Interface Hypothesis studies as well as studies done on L2 Japanese. In Chapter 3, I explore theories of overt pronouns, including the strong/weak pronoun distinction, the Position-of-Antecedent processing strategy (Carminati, 2002), and various anaphora hierarchies, as well as Gricean theories of anaphora. In Chapter 4, I present an overview of Japanese overt pronouns and offer a particularized conversational implicature (PCI) analysis of overt 3rd person pronouns in Japanese. In Chapter 5, I describe the research questions, methodology, and predictions of my experiment, with a summary of the results in Chapter 6. In Chapter 7, I analyze the results of the experiment in light of my analysis of Japanese 3rd person pronouns offered in Chapter 4.
CHAPTER 2
SECOND LANGUAGE ACQUISITION OF OVERT PRONOUNS

2.1 Background: The Interface Hypothesis

Many recent experiments on L2 overt pronoun use have focused on the Interface Hypothesis, which claims that near-native L2 and bilingual speakers diverge from monolingual speakers on constructions that involve interfaces between different language modules (i.e., syntax, semantics, etc.), as opposed to within a single language module (i.e., narrow syntax). In particular, external interfaces, such as the syntax-pragmatics\(^1\) interface, are considered to be a source of L1/L2 divergence due to increased processing strain (Sorace & Serratrice, 2009). These studies claim that overt pronoun use in null subject languages falls along the syntax-discourse interface, since a syntactic manipulation (presence or absence of a pronoun) affects interpretation on the discourse level; specifically, that the overt pronoun triggers switch-reference/topic-shift (from the subject to some other referent). Specifically, early studies theorized that differences between bilingual and monolingual groups stemmed from crosslinguistic influence involving a conflict between [-topic-shift]-bearing English overt pronouns and their [+topic-shift] bearing Italian/Romance counterparts (Sorace, 2000; Tsimpli et al., 2004; Sorace, 2005). Later studies have emphasized processing accounts and input exposure as possible (replacement or supplemental) explanations for this phenomenon (Sorace & Filiaci, 2006; Sorace, 2011). Crucially, differences between monolingual and bilingual groups are framed in terms of optionality, not grammaticality. That is, choosing a subject antecedent for an

\(^1\) The terms syntax-discourse and syntax-pragmatics interface are often used interchangeably in the literature, although the former is a subset of the latter.
overt pronoun is not ungrammatical for either monolinguals or bilinguals, but the latter tend to do so more often than the former in various experiments.

2.2 Previous Interface Hypothesis (IH) Studies

2.2.1 Childhood Bilingualism

Early work on syntax-pragmatics interface (often referred to as C-domain) phenomena in language acquisition emerged in the study of childhood bilinguals, with the hypothesis that overlap between two languages would result in crosslinguistic influence from a language with consistent input on a given C-domain feature (i.e., object drop in Dutch and German) to a language with inconsistent input for the same feature (i.e., object drop in French). Analyses of child corpus data seemed to bear out this hypothesis (Hulk & Muller, 2000; Muller & Hulk, 2001). However, later studies of older childhood bilinguals specifically testing overt pronoun usage have suggested processing factors as an alternate (or supplemental) explanation to transfer in monolingual and bilingual divergence. For example, Sorace & Serratrice (2009) found that in addition to English-Italian child bilinguals, Spanish-Italian child bilinguals also differed significantly from monolingual Italian controls in a task targeting overt pronoun use. Assuming that Italian and Spanish exhibit similar overt/null pronoun patterns, the authors suggest that differences may result from extra bilingual processing strains rather than from crosslinguistic influence.2 Furthermore, the role of input/exposure has been emphasized as a further complicating factor. For example, Argyri & Sorace (2007) found that Greek-dominant English-Greek bilingual children did not differ significantly from monolingual Greek adults on a

---

2 More recently, however, Filiaci, Sorace & Carreiras (2013) point out differences in overt pronouns between Spanish and Italian, arguing that crosslinguistic influence may play a role in monolingual and bilingual divergence after all, in addition to processing differences. We will return to this point in more detail in the next chapter.
judgment task involving the overt/null pronoun contrast, whereas English-dominant bilingual children did; Sorace & Serratrice (2009) found that English-Italian bilingual children (living in the UK) differed significantly from both Italian-English bilingual children (living in Italy) and Italian monolingual children, choosing an overt pronoun more often in topic-shift contexts.

2.2.2 Adult Bilingual Populations

Additional studies have found similar effects in adult bilingual populations, including L1 speakers of null-subject languages living in an environment where their non-null-subject L2 is the dominant language (referred to as “L1 attriters”) and near-native L2 speakers of null-subject languages. Tsimpli, Sorace, Heycock & Filiaci (2004) present an example of the former, testing L1-Italian near-native speakers of English with at least six years of residence in the UK on overt vs. null pronoun use in Italian. Participants were asked to identify the possible referents of the subject of the matrix clause in sentences with forward and backward anaphora, such as the Italian examples below, by selecting a picture corresponding to their interpretation of the sentence (Picture-Verification or PVT Task) (Tsimpli, et al., 2004, p. 266):

(1)  *L’anziana signora* saluta la *ragazza* quando *lei*/*pro* attraversa la strada.

the old woman greets the girl while she crosses the street

“The old woman greets the girl while she/pro crosses the street.”

(2)  *Quando lei*/*pro* attraversa la strada, *l’anziana signora* saluta la *ragazza*.

while she crosses the street the old woman greets the girl

“While she/pro crosses the street, the old woman greets the girl.”

Monolinguals tended to choose a topic-shift interpretation for overt pronouns in both forward- and backward-anaphora sentences; that is, they chose *the girl* or an external referent as the antecedent of *lei*. Bilinguals, on the other hand, had a tendency to choose the subject of the matrix clause (*the old woman*) as the antecedent of *lei*, in both forward- and backward-anaphora
sentences. Full results are reported below (Tsimpli et al., 2004, p. 267-273):

Table 1. Results of Tsimpli et al., 2004 (Forward Anaphora)

<table>
<thead>
<tr>
<th></th>
<th>Forward Anaphora</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Overt</td>
</tr>
<tr>
<td></td>
<td>Subject</td>
</tr>
<tr>
<td>Monolingual</td>
<td>8%</td>
</tr>
<tr>
<td>Bilingual</td>
<td>21%</td>
</tr>
</tbody>
</table>

Table 2. Results of Tsimpli et al., 2004 (Backward Anaphora)

<table>
<thead>
<tr>
<th></th>
<th>Backward Anaphora</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Overt</td>
</tr>
<tr>
<td></td>
<td>Subject</td>
</tr>
<tr>
<td>Monolingual</td>
<td>12%</td>
</tr>
<tr>
<td>Bilingual</td>
<td>~25%</td>
</tr>
</tbody>
</table>

Overall, the results support the IH for L1 attrition, with the bilingual group choosing subject antecedents for overt pronouns more often than monolinguals. In line with other early IH studies, the authors attribute monolingual and bilingual divergence to transfer of English interpretable pronoun features ([-topic-shift]) to the L1 Italian.

Other studies found similar results for L1-English, near-native L2 speakers of Italian (Sorace & Filiaci, 2006; Belletti, Bennati & Sorace, 2007). Sorace & Filiaci (2006) used the same methodology as Tsimpli, et al. (2004) and found largely similar results, with L1-English,

---

3 While there were no significant differences in the backward-anaphora/null condition, monolingual and bilingual groups differed significantly in the forward-anaphora/null condition, with the bilingual group choosing the non-topic-shift interpretation more often. Tsimpli et al. (2004) claim that bilinguals treat the subordinate clause as non-finite, like in English (p.273).

4 The authors do not report specific numbers for all of the percentages; those preceded by “~” are estimated from bar graphs.
near-native L2 Italian participants choosing the subject antecedent for the overt pronoun more often than monolingual Italian controls in both forward- and backward-anaphora sentences. Unlike in Tsimpli, et al. (2004), there were no significant differences between groups in either null condition (Sorace & Filiaci, 2006, p. 354):

**Table 3. Results of Sorace & Filiaci, 2006 (Forward Anaphora)**

<table>
<thead>
<tr>
<th></th>
<th>Forward Anaphora</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Overt</td>
<td>Null</td>
</tr>
<tr>
<td></td>
<td>Subject</td>
<td>Complement</td>
</tr>
<tr>
<td>L1</td>
<td>8%</td>
<td>82%</td>
</tr>
<tr>
<td>L2</td>
<td>27%</td>
<td>60%</td>
</tr>
</tbody>
</table>

**Table 4. Results of Sorace & Filiaci, 2006 (Backward Anaphora)**

<table>
<thead>
<tr>
<th></th>
<th>Backward Anaphora</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Overt</td>
<td>Null</td>
</tr>
<tr>
<td></td>
<td>Subject</td>
<td>Complement</td>
</tr>
<tr>
<td>L1</td>
<td>12%</td>
<td>24%</td>
</tr>
<tr>
<td>L2</td>
<td>47%</td>
<td>25%</td>
</tr>
</tbody>
</table>

Sorace & Filiaci (2006) conclude that the L2 participants had knowledge of both L1 narrow syntactic properties and L1 processing strategies (since they performed native-like in the null condition), but were more lax in their interpretation of overt pronoun. They attribute this non-native-like interpretation of overt pronouns to a lack of “the processing resources to integrate multiple sources of information consistently” (Sorace & Filiaci, 2006, p. 361), citing in particular the greater divergence in the backwards anaphora condition, which requires greater processing demand. Sorace & Filiaci (2006)’s results were largely replicated in Belletti, Bennati & Sorace (2007), including a storytelling task in which L2 speakers produced more overt pronouns than L1
controls, and a similar PVT task as in the two previously-mentioned studies with similar results. Belletti et al. (2007) suggest that the monolingual and bilingual speakers show “different instantiations of principles of economy;” that is, overt pronouns are more “costly” to monolinguals than to bilinguals, essentially a transfer argument (Belletti, et al., 2007, p. 682).

On the other hand, a study by Rothman (2009) testing intermediate and advanced L1-English/L2-Spanish speakers found that the most advanced learners did not differ significantly from monolinguals (although intermediate learners did) in (i) a pragmatic-context judgment task, in which participants rated sentences with null and overt pronouns in various contexts, and (ii) a pragmatic context translation task, in which participants were asked to translate a sentence from English into Spanish in contexts designed to elicit null vs. overt pronoun use in [+/− contrastive focus] situations. These results suggest, contrary to the original formulation of the IH, that near-native speakers are indeed able to acquire interpretable interface features (or native-like interpretation patterns, in more theory-neutral language). Later studies have also suggested that IH effects are not permanent, and that L2 speakers can ultimately attain native-like performance on and comprehension of other syntax-discourse interface structures, such as clitic left dislocation (Slabakova & Ivanov, 2011). While Rothman (2009)’s study casts some further doubts on a purely representational account of IH effects, it does not rule out a processing account, as it could be the case that untimed written tasks used in his study did not create the processing strains necessary to reveal the effects found in other IH studies.

2.2.3 Explaining Monolingual and Bilingual Divergence: Transfer, Processing, and Input

The experimental studies described above provide evidence to suggest that various bilingual populations diverge from their monolingual counterparts on the interpretation of overt pronouns in null argument languages, one of several structures that involve the syntax-
pragmatics interface. It is important to emphasize that these studies describe monolingual and bilingual *divergence*, not bilingual/L2 “incorrectness” or “ungrammaticality.” Sorace (2011, p. 3) refers to this divergence as “residual optionality” (on the part of near-native L2 speakers) and “emerging optionality” (on the part of L1 attriters), reflecting the fact that although monolingual and bilingual groups demonstrate different quantitative results, they show similar overall qualitative patterns (preferring a non-subject antecedent for overts more often than not).

Both representational and processing accounts have been offered to explain this “residual optionality,” with input/exposure often mentioned as a mitigating factor as well. The original representational account has somewhat fallen out of favor, at least as a sole explanatory factor, as bilinguals whose two languages are similar (i.e., Spanish/Italian) show the same effects as those whose languages differ (i.e. English/Italian) in terms of the same feature (null/overt pronoun alternation). The processing account currently claims that although syntax-pragmatic interface constructions are complex for both monolinguals and bilinguals, for the latter that complexity may be aggravated by other cognitive constraints. In particular, some studies have shown that executive control in bilinguals may be constrained by resource allocation issues due to the need to devote extra resources to the inhibition of the language currently not being used (Bialystock, 2009; Wilson, 2009; Sorace, 2011). Sorace (2011) argues that the difficulty of integration of information across syntax-pragmatics interface structures leads bilinguals to use a “heuristic” strategy of favoring redundant forms over ambiguous ones, thus resulting in maintain-reference interpretations for overt pronominals (Sorace, 2011, p. 31). The processing account is in line with Rothman (2009), which showed that with heavy context (providing additional cues beyond morphosyntax) and an untimed writing task (providing less processing pressure), advanced L2 participants performed similarly to monolingual native speakers.
Although several studies suggest that the overall quantity of target-language input may also play a role in monolingual and bilingual divergence (Argyri & Sorace, 2007; Sorace, 2005, Sorace & Serratrice, 2009), the exact mechanism through which quantity of input affects the language system has been only tentatively explored. Sorace (2011) suggests that priming/antipriming effects may result in increased activation of a given structure in the environmentally-dominant language at the expense of decreased activation of related structures in the non-dominant language (p. 22). Decreased input in one language also results in less opportunities to practice (both in terms of production and comprehension), which may keep speakers from “acquiring and maintaining an efficient syntactic system;” this general decrease in efficiency of the syntactic system leads to processing difficulties and therefore less available resources for complex syntax-pragmatics interface constructions (Sorace, 2005, p. 74). The childhood bilingualism studies in particular postulate that qualitative differences in the input from the so-called attrited adult population may shape the syntactic representations for children in the non-environmentally-dominant language (Argyri & Sorace, 2007; Sorace & Serratrice, 2009), leading the children to exhibit experimental results that diverge from their monolingual counterparts.  

2.2.4 Summary

In sum, a variety of experimental psycholinguistic studies have offered evidence that various bilingual populations diverge from their monolingual counterparts on overt pronoun use in null-subject languages. Originally, this phenomenon was attributed to a transfer account, in

---

5 Although none of the above-cited studies make the connection, sociolinguistic studies of language contact situations also describe changes in which the dominant language affects the non-dominant language of a minority speech community, with overt pronoun use in Spanish as a particular example (Otheguy & Zentella, 2012).
which speakers transferred an interpretable [-topic-shift] feature on overt pronouns from their non-null-subject language to their null-subject language. While syntax-pragmatics interface structures are still hypothesized to be sources of divergence for bilingual and monolingual speakers, explanatory factors beyond transfer of features (processing, input) have also been emphasized. The processing account (either as a replacement or supplement for the representational account) suggests that the effort put into inhibition of the other language results in exacerbated processing strain for bilinguals on complex syntax-pragmatic interface structures, and is supported by various arguments: (i) that monolingual/bilingual divergence appeared even when both languages were null-subject, (ii) that greater differences appear in backwards-anaphora conditions, which are more difficult to process than forwards-anaphora, (iii) that even those L2 speakers with non-null-subject L1s demonstrated knowledge of null-subject grammars, with native-like interpretations of the null pronoun, and (iv) that advanced L1-English/L2-Spanish participants in Rothman (2009) failed to show divergence when given a different task with less processing pressure. The role of input quantity and/or quality as another mitigating factor has also been offered, with long-term priming (of the environmentally-dominant language) and antipriming (of the environmentally non-dominant language) affecting overt pronoun interpretation.

2.3 Overt and Null Pronouns in L2 Japanese

Japanese, like the languages described above, also allows null arguments, so the question arises as to whether or not bilingual Japanese populations also diverge from monolinguals on overt pronoun interpretation as predicted under the Interface Hypothesis. Based on a handful of experimental studies, the answer is complicated: L2 participants tend not to diverge from L1 participants in these studies, but L1 participants do not show a consistent topic-shift/switch-
reference effect for overt pronouns in the first place. For example, Kanno (1997), although primarily interested in the Overt Pronoun Constraint (OPC)\(^6\), included sentences that fit the IH mold as control items, with an overt/null pronoun alternation in an embedded clause with a referential NP subject for the matrix clause (Kanno, 1997, p. 282):

(3)  \textit{Tanaka-san-wa raishuu \textit{kare-ga} \textit{o} Toukyou-e iku-to itteimashita yo.}  Tanaka-Mr.-TOP next.week he-NOM/ø Tokyo-to go-that said SFP

“Mr. Tanaka said that he/ø would go to Tokyo next week.”

Native Japanese and L1-English/L2-Japanese participants were given a pencil-and-paper survey in which they read each sentence and then were asked who did the action in the embedded clause.\(^7\) For the example above participants were asked “Who will go to Tokyo?” and were given the following choices: (a) Tanaka, or (b) someone else, with the option to circle both (a) and (b).

Results found that both L1 (N = 20) and L2 participants (N = 28; L1 English) found overt pronouns embedded under referential NPs to be ambiguous between a subject-coreferential and sentence-external interpretation. For the embedded 3\(^{rd}\) person pronoun kare, L1 participants permitted a matrix (subject-coreferential) antecedent 47\% of the time, while disallowing one the remaining 53\%; for L2 speakers the respective numbers were 42\% and 58\%.

However, note that Kanno (1997)’s analysis combines the “(a)-only” and “both (a) & (b)” options into a single category (“allows matrix-subject antecedent”), and contrasts that combined category with the “(b)-only” option (“does not allow matrix-subject antecedent”). A close look at the results shows that L1 participants were largely split between the “both (a) & (b) option” (both

\(^6\) The OPC hypotheses that overt pronouns in null subject languages cannot be bound by a quantifier; if \textit{Tanaka-san} in (3) were replaced with \textit{dare-ka} “someone”, \textit{kare} “he” should bias an external referent (not \textit{dare-ka}). This is what Kanno (1997) found for native speakers.

\(^7\) Kanno (1997)’s L2 participants had four semesters of in-class Japanese but had never been to Japan. They therefore probably do not fall under the near-native category, as is presupposed in most IH studies.
same-subject and external referent allowed) and the “(b)-only option” (only external referent allowed), whereas the L2 participants were split between the “(a)-only” and the “(b)-only” options (Kanno, 1997, p. 272-273):

Table 5. Results of Kanno, 1997

<table>
<thead>
<tr>
<th>Permits matrix subject antecedent</th>
<th>Does not permit matrix subject antecedent</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>(a)-only</em> (&quot;Tanaka&quot;)</td>
<td><em>(a) or (b)</em></td>
</tr>
<tr>
<td><em>(b)-only</em> (&quot;someone else&quot;)</td>
<td></td>
</tr>
<tr>
<td>L1</td>
<td>7%</td>
</tr>
<tr>
<td>L2</td>
<td>39%</td>
</tr>
</tbody>
</table>

That is, native speakers more often chose the option that reflects pure ambiguity (“both (a) and (b)”), with an additional bias towards one of those two interpretations (“(b)-only”), compared to the L2 participants, who alternated between one of two mutually-exclusive options. Overall, these results suggest that L1 participants interpret overt pronouns ambiguously, with a slight preference for an external referent, while L2 participants, although showing superficially-similar results to L1 participants, may just be guessing.

Okuma (2012) more closely replicated the IH paradigm with native Japanese (N = 14) and L1-English/L2-Japanese (N = 11) participants. Items included sentences with subordinate and matrix clauses (Okuma, 2012, p. 7):8

(4) *Okaasan-wa musume-ni kanojo-ga/o kooto-o kiru toki-ni kisu-o sita.*

mother-TOP daughter-DAT she-NOM/ø coat-ACC put.on when kiss-ACC did

“The mother kissed the daughter when she/ø put on the coat.”

Participants were asked to choose a picture corresponding to their interpretation of the sentence, with a choice of the mother, the daughter, and a third, external referent putting on the coat. L1 8

---

8 Okuma (2012)’s items included four different syntactic conditions in order to test word order and topic-marking. I report on the condition that most closely matches the previously-cited IH studies (in the forward anaphora condition, the only order that is allowed in Japanese).
participants had a matrix-subject antecedent preference for subordinate null subject pronouns (73% of responses), but no antecedent preferences one way or the other for overt subject pronouns (matrix-subject = 40%, indirect-object = 45%, external referent = 15%). L2 participants differed significantly from L1 participants only on null subjects, where they had a greater tendency than L1 participants to prefer a non-subject (topic-shift) antecedent (about 50% of the time). A production study found similar results for L1 participants, with no significant differences between L1 and L2 groups in any condition. In sum, Okuma (2012) provides no evidence that L2 Japanese participants have difficulty with the interpretation of overt pronouns; however, she also found no strong evidence for topic-shift in either the L1 or the L2. Once again, overt pronoun interpretation, while favoring more topic-shift responses than the null condition, seems to reflect ambiguity, with a near chance split on interpretation.

A pilot study by Yamada (2008; 2009), noting the rarity of overt 3rd person pronouns in Japanese, focuses on the NP/null alternation rather than the overt/null pronoun alternation. Participants included five native English and five native Korean learners of Japanese, all of whom had passed the highest level of the Japanese Language Proficiency Test (or equivalent), and five native-speaker controls. In a forced written-elicitation task, participants were shown a picture depicting a contextual situation along with a sentence describing the situation, then given a question sentence relating to the situation. The crucial items for overt form use were those in the Topic context (Yamada, 2009a, p. 62; Yamada, 2009b, p. 11-12):

---

9 Okuma (2012) notes that this divergence between groups on null subjects is unexpected, and suggests transfer or task effects. Based solely on the example item given in the text, there may be a lexical semantic issue at play, in that the embedded verb in the example (kiru “put on”) is non-causative in Japanese but can be causative in English. English speakers therefore, may have arrived at the interpretation “…while the mother put the coat on the daughter, [pro] kissed [pro],” in contrast to L1 speakers, who would only have available the structure “the mother, while [pro] put on [pro]’s coat, kissed her daughter.”
Results showed that while L1 Korean and L1 Japanese speakers overwhelmingly chose null forms in the topic-object condition (96% Korean; 88% Japanese), native English speakers were closely split on choosing a null form (60% of the time) or an overt form (40% of the time).\(^\text{10}\) Yamada (2008) chooses an interpretable-feature-based transfer account combined with shallow processing to explain the L1-English data (Yamada, 2008, p. 13). Processing factors may particularly exacerbate English participants’ responses, since there were no differences in the subject-topic condition in (5), in which participants must keep track of a single argument (the boy), as opposed to the object-topic condition in (6), in which participants must keep track of two arguments (the mother and the son).

\(^\text{10}\) The difference was not statistically significant, but this may have been due to the low number of participants.
In sum, Japanese studies relevant to the Interface Hypothesis reveal (i) pure ambiguity for overt pronoun interpretation by L1 participants, and (ii) some tentative differences between L1 and L2 speakers. Kanno (1997) found no strong antecedent preference for overt pronouns by native Japanese speakers; her L2 results look superficially similar to the L1 results, but may simply reflect guessing, rather than the principled judgment of ambiguity made by L1 speakers. Yamada (2008) suggests that processing strain may result in the over-realization of overt arguments by L2 speakers in production, although her study does not focus specifically on overt pronouns. Only Okuma (2012) provides evidence the L1 and L2 speakers interpret overt pronouns in a similar way, with both groups choosing an ambiguous interpretation rather than a clear antecedent preference. Since her interpretation task was forced choice, there is no way to tell whether this similarity masks similar findings to Kanno (1997)’s; that is, ambiguity for L1ers vs. random guessing for L2ers (although her production data suggests otherwise, since the groups performed similarly on that task).

The fact that L1 Japanese differs from Italian in the studies discussed above is somewhat unsurprising due to several differences between the two, with the distinction between Romance-like pro-drop languages and East Asian null argument languages long noted in the literature (Huang, 1984; 1989). Although Japanese allows null subjects, it differs from pro-drop languages in several key ways, including a lack of verbal agreement morphology for recovering person and number features, as well as fewer pronouns overall, partially due to frequent null objects and infrequent use of possessive pronouns, as well as more use of lexical NP forms. Furthermore, overt pronouns in Japanese display mixed lexical properties (Kuroda, 1965). I explore these differences in further detail in Chapter 4.
2.4 This Study

In this paper, I further explore the nature of overt 3rd person pronoun interpretation in Japanese, with the ultimate goal of comparing L1 and L2 interpretation patterns, as well as making connections to previous IH studies. In the next chapter, I examine in greater detail what it means when a language offers both null and overt pronouns as grammatical, commonly-used anaphors, exploring theories of strong vs. weak pronouns, anaphora hierarchies, and Gricean-based systems of anaphora. Then, in Chapter 4, I describe the properties of overt pronouns in Japanese (with a particular focus on 3rd person pronouns) and explore how they may or may not fit in with the theories of anaphora described in Chapter 3, ultimately suggesting a particularized conversational implicature analysis. I then argue that by providing contextual information specifically designed to encourage the topic-shift reading, I will be able to elicit a robust (more than chance) tendency for topic-shift in L1 Japanese participants in contrast to previous experiments. It then remains to be seen whether L2 participants will pattern with the L1 speakers or not using the same test items.
CHAPTER 3

OVERT AND NULL PRONOUNS IN THEORY

3.1 Introduction

This study aims to compare the interpretation of overt pronouns by native and advanced L2 speakers of Japanese. It was inspired by a series of second-language acquisition (SLA) studies on overt pronoun interpretation in null subject languages testing the Interface Hypothesis (IH), which claims that certain bilingual populations (including near-native L2 speakers) diverge from native monolinguals on language forms that involve the interface between two language modules, in particular, syntax and discourse (see Sorace, 2011 for an overview). One key linguistic property underlying many of these studies is the association of topic-shift (or switch-reference) with overt pronouns in Romance-style null-subject languages such as Italian, Spanish, and Greek. In this chapter, I review previous literature on topic-shift/switch-reference in Romance languages, explore anaphora hierarchies as an explanation of this phenomenon, and describe Neo-Gricean theories of anaphora that, as I will argue in Chapter 4, can model a crucial difference between pronoun use in Japanese vs. various other null argument languages.

3.2 Behavior of Overt Pronouns in Null Argument Languages

3.2.1 Evidence from Psycholinguistic Studies

Early SLA studies on the Interface Hypothesis postulated an interpretable [+ topic-shift] feature on overt pronouns in Romance-like languages (typically, Italian and sometimes Greek), sitting at the interface between syntax and discourse (or pragmatics) (Sorace, 2000; Tsimpli et al., 2004; among others). As described in Chapter 2, typical test items in this paradigm include those like (7) and (8) below, with a critical alternation between overt and null pronouns in a subordinate clause, in sentences with both forward and backward anaphora. Under the
interpretable feature approach, since the overt pronoun bears a [+topic-shift] feature, participants are likely to report that she refers to the girl rather than the old woman (i.e., shift from the subject of the matrix clause to the direct object), whereas the null pronoun (pro), which bears no such topic-shift feature, biases an interpretation in which the old woman does both actions (Tsimpli, et al., 2004, p. 266):

(7)  *L’anziana signora* saluta la *ragazza* quando *lei*/*pro* attraversa la strada.
   The old woman greets the girl while she crosses the street
   “The old woman greets the girl while she/pro crosses the street.”

(8)  *Quando lei*/*pro* attraversa la strada, *l’anziana signora* saluta la *ragazza*.
   While she/pro crosses the street, the old woman greets the girl
   “While she/pro crosses the street, the old woman greets the girl.”

Participants in these studies were asked to choose the picture that best represented each sentence, with a choice of (i) the old woman crossing the street, (ii) the girl crossing the street, (iii) a third, unmentioned, woman crossing the street. Results showed that monolingual Italian participants consistently chose a non-subject referent (either the mentioned non-subject complement or the external referent) in the overt pronoun variant, roughly 90% of the time in both forward and backward anaphora conditions (Sorace & Filiaci, 2006, p. 354):

**Table 6. Results of Sorace & Filiaci, 2006 (Native Italian Speakers)**

<table>
<thead>
<tr>
<th></th>
<th>Forward Anaphora</th>
<th>Backward Anaphora</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Subject</td>
<td>Complement</td>
</tr>
<tr>
<td>Null</td>
<td>51%</td>
<td>44%</td>
</tr>
<tr>
<td>Overt</td>
<td>8%</td>
<td>82%</td>
</tr>
</tbody>
</table>

Belletti et al. (2007) largely replicated these results for their L1 Italian controls with similar items. Rothman (2009) found similar results for L1 Spanish using different tasks, including a written production task.
Additional evidence is found in a sentence-processing study of native Italian speakers by Carminati (2002), which theorizes a processing strategy called the Position of Antecedent Hypothesis (PAH). The PAH is a purely structural constraint on processing, stating that Italian speakers prefer an antecedent in SpecIP for null pronouns, whereas they prefer an antecedent in some other (lower) position for overt pronouns.\(^{11}\) Carminati (2002) presents the results from several experiments testing the PAH under various conditions, including test items with both intrasentential and extrasentential anaphora, sentences with non-canonical subjects, different types of phi-features, and quantifier antecedents. Results largely supported predictions for both null and overt pronouns, although Carminati (2002) notes that overt pronouns seemed more sensitive to contextual factors than nulls. Specifically, interpretation of the overt straightforwardly follows structural factors, choosing referent lower than Spec-IP, in purely ambiguous contexts, but follows structural predictions less robustly in unambiguous contexts (p. 303-304). In other words, participants tended to prefer non-subject antecedents for overt pronouns (topic-shift), but this preference was more sensitive to contextual factors than the subject-antecedent preference of the null. Overall, the experimental data in these SLA and sentence-processing studies confirm the idea that native monolinguals of these languages tend to interpret overt pronouns as indicating some sort of topic-shift, a tendency that can then be compared to the interpretation of the same sentences by various bilingual populations.

3.2.2 Evidence from Variationist Studies

While the studies in 3.2.1 show the association between overt pronouns and topic-shift interpretation in experimental settings, several variationist studies have explored overt pronoun

---

\(^{11}\) The IH literature later borrowed the PAH in developing a processing account of L1/L2 divergence (as opposed to the topic-shift feature transfer analysis).
use in null subject languages using corpus data. In these studies, both linguistic and social variables are used to statistically predict the appearance of an overt pronoun in corpus data, with the former including both discourse factors (such as switch-reference/topic-shift) and morphosyntactic factors (such as tense) and the latter including variation among speaker groups (such as regional differences). Switch-reference has long been described as an important predictor of the appearance of an overt pronoun in such studies. Cameron (1995), examining interview transcripts from Spanish speakers in San Juan (N = 10), found that in same-reference contexts, the trigger was an overt pronoun 31% of the time, but in switch-reference contexts 57% of the time, a difference of 26 percentage points. Transcripts of speakers from Madrid (N = 10) showed a lower rate of overt pronoun use overall, but with the same pattern: 11% in same-reference and 30% in switch-reference contexts. Cameron (1995) found that a strong similarity between the groups, with probabilities (VARBRUL weights derived from logistic regression coefficients for each factor) for switch reference at .64 (San Juan) and .65 (Madrid) (p. 10).

Cameron also notes that similar results were found in previous studies for Spanish speakers in Venezuela (Bentivoglio, 1980) and Los Angeles (Silva-Corvalan, 1989).

Otheguy, Zentella & Livert (2007) investigated ten possible linguistic factors predicting the presence of an overt pronoun in interview data from Spanish speakers in New York City: genre of discourse, person and number of verb, verb tense, reflexivity of verb, specificity of reference, switch-reference (discourse connection), lexical content of verb, clause type, whether verb is in a set phrase, and section of interview. The linguistic factors were put into a logistic regression and ranked by Wald statistic. For newly-arrived speakers (over age 16 and in NYC for five years or less), switch-reference was ranked second after verb person/number features for both Caribbean (N = 19) and Mainland Latin America (N = 20) groups (p. 789). These findings
show that even when multiple linguistic factors are taken into account, switch-reference is the second-most powerful predictor of where an overt pronoun will appear, across two different regional groups.

3.3 Interpreting Overt and Null Pronouns

Evidence from both psycholinguistic and corpus studies, therefore, reflects a strong association between overt pronoun use and topic-shift/switch-reference in null subject languages. In this section we further explore some theories of anaphora that attempt to explain this tendency, including the Structural Deficiency Hypothesis of Cardinaletti & Starke (1999), upon which the IH interpretable feature account was based, various anaphora hierarchies (Givon, 1983; Ariel, 1990; Gundel et al., 1993), and Neo-Gricean anaphora systems (Huang, 1994; Levinson, 1987; 2000).

3.3.1 Strong and Weak Pronouns

The early IH interpretable feature analysis drew on theoretical work by Cardinaletti & Starke (1999), which postulates a classification of pronouns as “strong” vs. “deficient,” with the latter further broken down into “weak” and “clitic” (Sorace, 2000; Cardinaletti & Starke, 1999). This analysis claims that pronouns are more or less deficient on several different modules (phonological, morphological, syntactic, and semantic); the less structure in each of these categories, the weaker the pronoun. Under this analysis, in Italian, null pronouns are deficient (“weak”) compared to overt pronouns (“strong”), as manifested in their phonological and morphosyntactic structure. Sorace (2000) makes this analysis the basis for the representational account in the IH studies by further arguing that one feature that makes overt pronouns semantically stronger than their weak null counterparts is a [+topic-shift] interpretable feature.
Of course, this [topic-shift] feature analysis needs is mechanism that accounts for the fact that topic-shift triggered by the overt pronoun represents a general pragmatic tendency, not a categorical grammatical distinction (i.e., a non-topic-shift overt interpretation runs contra to pragmatic tendencies, not syntactic grammaticality). Sorace (2006) postulates more optionality/“graded” grammaticality effects for interface structures vs. non-interface structures even in monolingual speakers, with monolingual and bilingual populations sometimes diverging the degree to which this optionality manifests itself. In particular, the complexity of these interface structures are claimed to cause increased processing difficulty as speakers must integrate different kinds of knowledge (from the interfacing modules at hand; syntax and pragmatics in the case of overt pronoun interpretation). As such, later IH studies have shifted focus to processing accounts, such as Carminati (2002)’s PAH, which, as a processing strategy naturally predicts tendencies rather than categorical distinctions. Carminati (2002) cites anaphora hierarchies such as that of Ariel (1990) as theoretical background in her study. Below, we will explore these anaphora hierarchies before examining Gricean systems of anaphora that can be used to integrate these hierarchies into a wider pragmatic system (Gundel et al., 1993; Huang, 1994; Levinson, 2000).

3.3.2 Anaphora Hierarchies

There have been various anaphora hierarchies put forth, including Givon’s Topic-Continuity model (Givon, 1983), Ariel’s Accessibility Theory (Ariel, 1990), and Gundel et al.’s Givenness Hierarchy (Gundel et al., 1993). These models rank different kinds of anaphors on a scale based on the kind of antecedent each is likely to retrieve. At one end is usually a local antecedent (or topic, or subject, or other salient referent), retrieved by anaphors with little structure, such as nulls or agreement morphology. On the other end are new referents or other
similarly difficult-to-access antecedents, usually retrieved by a lexical NP. These models map referring expressions to antecedents, and, as, such, have two crucial components: a ranking of (and/or a mechanism for ranking) anaphors/referring expressions (such as null pronoun, overt pronoun, lexical noun, etc.) and some kind of antecedent-saliency scale (based on topic-hood, linear distance, grammatical position, and/or other factors). The exact nature of each of these components varies slightly in each model.

Givon (1983)’s model involves a scale that ranks antecedents from most to least continuous/accessible topic, with different referential expressions ranked on this scale from zero anaphora (most continuous/accessible) to referential indefinite NPs (least continuous/accessible). In terms of antecedent saliency, Givon (1983) focuses on “topic-continuity,” which, in the corpus studies testing his scale on various languages in the same volume, is assessed by three criteria: (1) referential distance (number of clauses back to the previous mention of the same referent), (2) potential ambiguity (number of competing referents in the previous 1-5 clauses), and (3) persistence (number of clauses for which a referent continues uninterruptedly as topic). Locality therefore is the major constraint on antecedent saliency, although Givon (1983) allows for the influence of other factors, such as grammatical position (subject, object, etc.). As for the ranking of referring expressions, Givon’s original list (inspired by English) appears below (Givon, 1983, p. 17):

Givon’s Topic-Continuity Hierarchy
most continuous/accessible
  zero anaphora
  unstressed/bound pronouns or grammatical agreement
  stressed/independent pronouns
  R-dislocated DEF-NPs

Cardinaletti & Starke (1999)’s weak/strong pronoun distinction posits a mechanism for ranking pronominal forms that is not inconsistent with these anaphora hierarchies.
neutral-ordered DEF-NPs
L-dislocated DEF-NPs
Y-moved NPs (“contrastive topicalization”)
cleft/focus constructions
referential indefinite NPs
most discontinuous/inaccessible topic

Under this analysis, a stressed pronoun (analogous to Cardinaletti & Starke (1999)’s strong pronoun and Sorace (2000)’s overt pronoun) is used for a more discontinuous topic than zero anaphora or grammatical agreement (analogous to Cardinaletti & Starke (1999)’s deficient pronouns or Sorace (2000)’s null pronouns). Givon’s hierarchy, therefore, also predicts that an overt pronoun in a null-subject language will tend to be used for and/or interpreted as an indication of switch-reference or topic-shift more often than a null form (although less often than a referential NP).

Although the above scale is often cited as-is, it is important to note that Givon (1983) actually breaks down this hierarchy into a series of smaller ones, based on factors like phonological size, word order, and semantic role/syntactic case. For example, the scale cited above can be divided into a smaller hierarchy in which referring expressions are ranked based on phonological size, and another where they are ranked based on word order. The phonological hierarchy is based on an iconicity principle, claiming that more discontinuous topics require more phonological material to retrieve (Givon, 1983, p. 18):

Givon (1983)’s Phonological-Size-Based Hierarchy
more continuous/accessible topics
  zero anaphora
  unstressed/bound pronouns (“agreement”)
  stressed/independent pronouns
  full NPs
more discontinuous/inaccessible topics

The word order scale, on the other hand, ranks referring expressions by the various dislocation constructions they may appear in, such as the following, ranked from most to least accessible:
Right-dislocation > neutral word order > Left-dislocation (Givon, 1983, p. 19). It is also noteworthy that Givon (1983) deferred from making any predictions about the interaction of topic continuity with what he calls morphological “subsystems,” including definite/indefinite markers and demonstratives, which later scales do incorporate into their ranking of referring expressions (Givon, 1983, p. 35). In sum, in Givon (1983)’s hierarchy, the two components are set as follows, although Givon (1983) leaves much room for flexibility: (i) an antecedent saliency scale based on “topic-continuity,” which largely (although not completely) derives from linear distance, and (ii) a ranking of anaphors that can be broken down into subscales, based on factors like word order and phonological weight.

Ariel (1990) offers a slightly different anaphora scale, with a slightly shifted focus on referent accessibility rather than topic-continuity (although with much overlap). She proposes four criteria defining antecedent accessibility: (1) distance, (2) competition, (3) saliency, and (4) unity. The first two are quite similar to Givon (1983)’s first and second criteria, the third captures “mainly” whether or not a referent is a topic, and the fourth whether the anaphor is thematically connected to the referent (Ariel, 1990, p. 29). In her ranking of anaphors, null pronouns (“zeros”) remain most likely to indicate high accessibility, with unstressed and stressed pronouns higher up the list. Ariel (1990) includes three criteria for ranking anaphors: Informativity, Rigidity, and Attenuation. The first is roughly equivalent to semantic content and the last to phonological size, with Rigidity more or less a semantic criterion as well (how likely the anaphor is to retrieve the antecedent unambiguously, i.e., a proper name compared to a common noun). One version of the Accessibility Hierarchy appears below (Ariel, 1991, p. 449):

*Ariel’s Accessibility Hierarchy*

LOW ACCESSABILITY

Full name + Modifier
Full name
Long definite description
Short definite description
Last name
First name
Distal demonstrative (+Modifier)
Proximal demonstrative (+Modifier)
Stressed pronouns + Gesture
Stressed pronouns
Unstressed pronouns
Zeros

HIGH ACCESSABILITY

Note that Ariel (1991)’s hierarchy, unlike Givon (1983)’s breaks down the category “full NP” into six sub-categories, including proper names, NPs with different degrees of modification, and adds demonstratives.

Gundel et al. (1993)’s Givenness Hierarchy is similar to the hierarchies of Givon (1983) and Ariel (1990), with referential expressions ranked along a discrete scale according to cognitive status of the antecedent, essentially filling in more detail on the points between “low” and “high” referent saliency found on the extremes of the previous scales. Below is the Givenness Hierarchy, with examples for English (Gundel et al., 1993, p. 275):

*Givenness Hierarchy*

in focus > activated > familiar > uniquely identifiable > referential > type identifiable

\{it\} > \{that, this, this N\} > \{that N\} > \{the N\} > \{indefinite this N\} > \{a N\}

The “in focus” end of the scale represents referents that are currently the main topic of discussion, with “activated” referents recently mentioned but not currently the focus of attention; both are retrieved from short-term memory. On the other end are “type identifiable” referents, where the hearer must summon the concept of the “type” of the referent from long-term memory, for example, to the concept of the type *dog* upon hearing “a dog.” The left end of the scale, therefore, includes referents already present in the conversation (in short-term memory), whereas the right end of the scale includes referents brought into the conversation via shared world
knowledge (long-term memory). Gundel et al. (1983) set items on their scale partially from theoretically-based predictions, and largely (although not perfectly) confirm these predictions with an analysis of corpus data. Observing the results of their corpus study, the authors note that anaphors with least phonetic and semantic content are associated with the in focus or activated status, that is, the statuses associated with the most recently-uttered (and therefore more accessible) referents. This is in line with both Givon (1983)’s and Ariel (1990)’s criteria for ranking anaphors.

Gundel et al. (1993) argue for different instantiations of the scale in different languages, noting that while languages tend to have distinctions for at least the two highest categories (in focus/activated), they may not have separate anaphors or anaphoric constructions for the lower end of the scale; this is possible and even likely since higher statuses entail all lower statuses. The authors give various instantiations of the hierarchy for different languages, with Spanish having a null pronoun or unstressed pronoun in the “in focus” category, and a stressed pronoun in the “activated” category, capturing the intuition that pronouns (as opposed to lexical nouns) usually refer, if not to the current topic, then to some relatively recent entity in the discourse. According to the hierarchy, since the stressed pronoun in Spanish is less preferred for the topic that is currently central to the conversation than the null or unstressed pronoun, use of a stressed pronoun may indicate a shift in topic, but doesn’t necessarily have to.\(^\text{13}\) Gundel et al. (1993) specifically discuss how their model works with Grice’s Maxim of Quantity in order to lead to interpretation effects for different anaphoric expressions on the scale, although their analysis is

\(^{13}\text{Although Gundel et al. separate “stressed” and “unstressed” Spanish pronouns on their scale, they group the two together in their corpus analysis later in the same paper since they were relying on written transcriptions.}\)
based on Grice’s original work (Grice, 1975), as opposed to more recently-developed Neo-Gricean theories of anaphora, which we will explore in Section 3.5 below.

3.4 Anaphora Hierarchies and Linguistic Variation

3.4.1 Variation in Anaphor Rankings

All three of the anaphora scales discussed in the previous section (Givon (1983), Ariel (1990), Gundel et al. (1993)) aim for universality, usually by claiming that the general pattern (two ends of the scale with some sort of ranking of forms in between) holds for all languages even if not all languages instantiate the same subset of encodings. This leads to the question of exactly how and why languages vary in the types of anaphors that they instantiate. All three models allow that languages vary based on the number of points on the scale they actually encode. Givon (1983) is part of a volume that provides corpus data from nine languages; Gundel et al. (1993) from five. Of these five, Gundel et al. (1993) note that English encodes all six cognitive statuses, Spanish five, and Russian, Japanese, and Chinese four. Japanese and Russian, for example, lack determiners, and therefore the three rightmost antecedent categories (uniquely identifiable, referential, and type identifiable) are all retrieved by the bare noun.

In addition to whether or not certain forms exist in a given language, there is also the problem of how frequently various forms occur. Gundel et al. (1993) leave open the question of why languages should have different frequencies of zero pronominals, as is evident in their corpus data. While they do not consider null pronouns for English, the rates of null tokens divided by all anaphor tokens for the other languages vary quite a bit: Chinese-11%, Japanese-24%, Spanish-6%, Russian-19%; the results are even more striking when only “in-focus” tokens (the highest status, where nulls are predicted to appear) are considered: Chinese-28%, Japanese-70%, Spanish-19%, Russian-36%. Givon (1983) also notes this frequency issue: he lumps
together English with Ute as languages that have both nulls and unstressed pronouns, but points out that the relative frequency of nulls and unstressed pronouns in each language differs, with English more likely to have unstressed pronouns for subjects and Ute more likely to have nulls. In a separate chapter on English spoken corpus data, he suggests that in addition to strict distance criteria (they must be within one clause of their antecedent), nulls in English must maintain strict overall thematic/action continuity, to a greater degree than unstressed pronouns (Givon, 1983a).

Ariel (1990) argues that the exact nature of the anaphora rankings varies among languages in two ways: (i) disallowing a form (marking it as ungrammatical) or (ii) restricting the distribution of a form, either grammatically or pragmatically. English, for example, does allow null subjects (in non-finite clauses, as subjects of conjoined clauses, etc.), but they are much more restricted than their counterparts in Japanese (where, instead, overt pronouns are more restricted). Ariel (1990) specifically points out that while the relative ordering of anaphors should hold cross-linguistically (i.e., nulls should not consistently retrieve a less-accessible antecedents than overt pronouns), high accessibility markers (nulls, pronouns) are “hardest to correlate cross-linguistically,” as they tend to be “spaced out differently on the Accessibility scale” from language to language (p. 65). On the other hand, markers on the lower side of the scale may simply be present or absent in a given system (i.e., definite/indefinite morphology associated with the NP), rather than overlapping in complicated ways, as on the higher end of the scale.

In sum, while these scales and the system for ranking the anaphors within them are claimed to be universal, the actual instantiation of the anaphors may vary from language to language. In particular, the higher an anaphor is on the scale (the more accessible of a referent it retrieves), the less it maps neatly from one language another; highly-accessible anaphor types
also tend to vary in (token) frequency from language to language as well. The frequency of nulls vs. unstressed pronouns is particularly tricky: many languages have both options available, but tend to use one more than the other. However, even though the distribution of each of these anaphors varies, the overall scale still maintains; for example, in English, where nulls are infrequent, they are nevertheless even more constrained than unstressed pronouns in maintaining topic-continuity.

3.4.2 Variation in Antecedent Saliency

The three anaphora systems described above don’t go into much detail about how the various constraints on antecedent saliency (accessibility, etc.) actually combine in order to influence the interpretation of individual anaphors. With linear distance, topic-hood, syntactic position, and other factors all postulated to constrain antecedent saliency, what role does each of these individual factors play, or are they all combined to make some sort of composite score for each possible antecedent? Kaiser & Trueswell (2008) offer a foray into this problem based on their experimental work on demonstratives and pronouns in Finnish. Results of a sentence-completion task and eye-tracking study suggest that participants tended to interpret the pronoun *han* as referring to previous subjects, while interpreting the demonstrative *tama* as referring to postverbal referents that are also discourse-new. In other words, when interpreting the pronoun, participants are sensitive to syntactic constraints on the antecedent, but when interpreting the demonstrative, they are sensitive to both syntactic and discourse factors. The authors argue for what they call the Form-Specific Multiple Constraints (FSMC) Hypothesis, which claims that each anaphor in a language has “its own set of weighted constraints that guide its interpretation” (Kaiser & Trueswell, 2008, p. 739); in other words, each anaphor interacts with a different subset of antecedent constraints. They note that the FSMC approach is compatible with scales such as
Ariel (1990)’s, as long as the constraints on antecedents aren’t taken to be equally weighted for all anaphors. They suggest that the FSMC is particularly useful for teasing apart anaphors that otherwise show a great deal of overlap in use, such as the pronoun and demonstrative in Finnish.

While Kaiser & Trueswell (2008) shed light on the properties of anaphors within a language, Filiaci et al. (2013) use the lens of FSMC to postulate differences between languages. They test Carminati (2002)’s Position-of-Antecedent Hypothesis (PAH) (a structural constraint on antecedent salience claiming nulls retrieve a referent in Spec-IP, overts a referent in a lower position, described in section 3.2.1) on native speakers of Spanish and Italian. Participants read sentences (phrase-by-phrase, self-paced) whose semantics forced coreference of a null or overt pronoun with a subject or object antecedent; participants then answered comprehension questions. Results showed no difference between the groups in processing nulls, in that both showed processing costs (longer reaction-times) when forced into a non-subject coreferential interpretation. However, when processing overt pronouns, Spanish participants did not demonstrate the same processing cost for subject antecedents that Italian participants did (measured in reaction time). For example, Spanish participants were more accurate than Italian participants in comprehension questions when the antecedent for an overt pronoun was a subject; Italian participants were slower when the antecedent for an overt was in subject position. The authors conclude that interpretation of Italian overt pronouns is more sensitive to the syntactic position of the antecedent than in Spanish; they suggest linear distance of the antecedent as a more powerful predictor for interpretation of overts in Spanish. The authors note that their results could be seen as evidence that the FSMC approach applies to forms across languages as well as within, or that pronouns in Italian and Spanish may be “two different types of lexical item” (p. 17). Specifically, they postulate that Spanish overt pronouns are weak and Italian overts are
strong, following Cardinaletti & Starke (1999); this idea is also compatible with Ariel (1990)’s observation that anaphors towards the top of the Accessibility scale are “spaced out” differently in different languages. This analysis also suggests that differences in overt pronoun interpretation found in Spanish-Italian bilinguals (Sorace & Serratrice, 2009), used to argue for a processing account, may reflect representational differences after all.

3.5 Neo-Gricean Theories of Anaphora

3.5.1 Background: Maxim of Quantity and Anaphora

Languages that commonly use both null and overt pronouns naturally invite a Gricean analysis, since each form is often found in places where the other could also appear. This brings to mind Grice’s Maxim of Quantity, which states:

(1) Make your contribution [to the conversation] as informative as is required (for the purposes of the exchange).

(2) Do not make your contribution more informative than is required. (Grice, 1975, p. 45).

Using an overt pronoun in a situation where a null could be used, therefore would constitute a flouting of the Maxim of Quantity, resulting in some sort of conversational implicature. This implicature could very well be a shift from a prominent to a less-prominent antecedent. In fact, this type of implicature serves as good candidate for the (or a) pragmatic mechanism by which the referring expressions on an anaphora scale map onto various types of antecedent.

Indeed, Gundel et al. (1993) incorporate Grice’s Maxim of Quantity into their system, although in a slightly roundabout way from what I described above. Since in their system higher cognitive statuses entail all lower statuses, they argue that the Maxim of Quantity provides further constraints on which anaphoric form is actually chosen to represent a given status. For
example, for the highest rank on their scale (“in focus”), speakers could technically choose any type of anaphor on the scale. However, the Maxim of Quantity leads speakers to choose the highest ranked anaphor, a null or overt pronoun, depending on the language. Ariel (1990) cites Grice (1975) in her analysis as well, but largely frames the pragmatic discussion of her system around Sperber & Wilson (1986)’s Relevance Theory, which builds on Grice (1975)’s Maxim of Relevance. Givon (1983) does not mention Grice, but his phonological iconicity scale fits in well with a Gricean analysis based on the Maxim of Quantity.

3.5.2 Particularized and Generalized Conversational Implicatures

One of the most fleshed-out Gricean systems of anaphora in a null argument language is Huang (1994)’s analysis of Mandarin Chinese. Huang’s work builds on a Neo-Gricean system developed by Levinson (1987, 2000), which reinterprets Grice (1975)’s Quantity and Manner Maxims into three “heuristic” principles: the Q[uantity]-, I[nformativeness]-, and M[anner]-principles. The Q-principle derives from Part (1) of Grice (1975)’s Maxim of Quantity, the I-principle from Part (2) of the Maxim of Quantity, and the M-principle from the Maxim of Manner.14 Crucially, Levinson and Huang’s work makes a distinction between particularized and generalized conversational implicatures, with their analyses of anaphora falling into the latter category. This distinction can be traced back to Grice (1975), who, towards the end of his paper, notes that most of his discussion of conversational implicatures in the article up to that point had been framed in terms of particularized conversational implicatures (PCIs). He describes PCIs as “cases in which an implicature is carried by saying that p on a particular occasion in virtue of special features of the context, cases in which there is no room for the idea

14 The Maxim of Manner consists of a supermaxim: “Be perspicuous” and four submaxims: “1. Avoid obscurity of expression. (2) Avoid ambiguity. (3) Be brief. (4) Be orderly. (Grice, 1975, p. 46).”
that an implicature of this sort is NORMALLY carried by saying p” (p. 56). He contrasts these particularized conversational implicatures with “generalized” conversational implicatures, in which “the use of a certain form or words in an utterance would normally (in the ABSENCE of special circumstances) carry such-and-such an implicature” (p. 56). Grice gives the example of the indefinite article “a” in English, noting that it usually implicates (via a GCI) a lack of ownership. For example, in (9) below, the hearer does not expect X to own the house or the tortoise (Grice, 1975, p. 56):

(9) X went into a house yesterday and found a tortoise inside the front door.

This implicature, however, may occasionally fail to arise, as in (Grice, 1975, p. 56):

(10) I have been sitting in a car all morning.

Levinson (2000) offers the following re-wording of Grice’s original argument (p. 16):

*The distinction between PCIs and GCIs*

a) An implicature i from utterance U is *particularized* iff U implicates i only by virtue of specific contextual assumptions that would not invariably or even normally obtain

b) An implicature i is *generalized* iff U implicates i unless there are unusual specific contextual assumptions that defeat it.

In other words, PCIs are calculated from scratch from the discourse context in the manner most stereotypically associated with Grice’s classic conversational implicatures (Levinson 2000) refers to them as “nonce” implicatures (p. 22)). The example of using an excuse (“I have some errands to run”) to decline an invitation, for example, is a PCI via the Maxim of Relevance.15

This kind of implicature is produced and interpreted uniquely in a particular discourse situation, as opposed to the GCI on the indefinite article, which automatically arises whenever the article is used, but may then be cancelled. Levinson (2000) theorizes that GCIs exist at a level between

---

15 Levinson (2000) does not have much to say about the Maxims of Quality and Relevance, as he claims they largely trigger PCIs, not GCIs, and the latter are his primary interest.
sentence meaning (akin to semantic truth-conditions) and speaker meaning (or utterance-token meaning, where PCIs come into play), which he designates “utterance-type meaning” (p. 22). This layer involves inferences based on “general expectations about how languages is normally used” rather than “computations about speaker-intentions (Levinson, 2000, p. 22-23). For indefinite article example, speakers do not have to actively calculate the GCI from context every time they encounter an indefinite article; the implicature instead arises by default. It can be cancelled by context, but context is not necessary to create the implicature in the first place.

3.5.3 A GCI Analysis of Anaphora

Huang (1994)’s pragmatic theory of anaphora, described largely through the lens of Mandarin Chinese data, incorporates a ranking of zero anaphora, pronouns, reflexives, and names/lexical NPs, and consists of two main parts: (a) Interpretation Principles, and (b) Consistency Constraints. The former describe four types of anaphors and the generalized conversational implicatures associated with each; the latter provide various mechanisms by which these implicatures can be cancelled. The Interpretation Principles cover zero anaphora (nulls), pronouns, reflexives, and lexical NPs, as follows (Huang, 1994, p. 145):

“Assuming that a reflexive is necessarily referentially dependent, and a pronoun and a zero anaphor are optionally but preferably referentially dependent:

i. the use of a zero anaphor will I-implicate a local coreferential interpretation

ii. the use of a pronoun will I-implicate a local coreferential interpretation, unless the pronoun is used where a zero anaphor could occur, in which case, the use of the pronoun will M-implicate the complement of the I-implicature associated with the use of the zero anaphor
iii. the use of a reflexive will I-implicate a local coreferential interpretation, unless the
reflexive is used where a pronoun or zero anaphor could occur, in which case, the use of
the reflexive will M-implicate the complement of the I-implicature associated with the use
of the pronoun or the zero anaphor, in terms of either reference or expectedness; and
iv. the use of a name or a lexical NP where a pronoun or a zero anaphor could occur will M-
implicate the complement of the I-implicature associated with the use of the pronoun or
the zero anaphor, in terms of either reference or expectedness.”

In other words, if a given anaphor is used where a “simpler” form could have been used
(pronoun for null, reflexive for pronoun), the M-principle (akin the Maxim of Manner) will
implicate the opposite of whatever that “simpler” form usually implicates. In the case of the
overt/null distinction, if an overt is used where a null could be used, since nulls normally indicate
local coreference, speakers assume non-local coreference. That is, the system of M-implicatures
built into Huang (1994)’s analysis is consistent with anaphora hierarchies that rank different
forms. It is also important to note nulls (zero anaphora) form a kind of baseline against which all
other anaphors are ultimately compared—nulls are associated with an I-implicature alone, with
no M-implicature based on contrast with another anaphor. The fact that they are pragmatically
constrained by one fewer implicature type is perhaps reflected in the fact that interpretation of
nulls is often seen as more “syntactically” or “structurally” constrained than that of overts, as in
Carminati (2002)’s PAH study.

The Consistency Constraints provide mechanisms through which the implicatures
described in the Interpretation Principles can be cancelled, and include (Huang, 1994, p. 145):

i. the Disjoint-Reference Presumption

ii. information saliency
iii. general implicature constraints

The Disjoint Reference Presumption, which Huang quotes from Farmer & Harnish (1987), states that “The arguments of a predicate are meant to be disjoint, unless marked otherwise” (in Huang, 1994, p. 129); it is largely used in the differentiation of pronouns and reflexives (and therefore of less interest here). Information saliency, analogous to some of the antecedent saliency factors in Ariel (1990) and Givon (1983), includes an antecedent hierarchy along the lines of topic > subject > object; saliency factors may create additional implicatures that override the GCIAs described in the Interpretation Principles. General implicature constraints include background assumptions, speaker intent, and semantic entailments, which again may cancel the GCIAs formed by the Interpretation Principles.

Huang (2000) argues that his system is complementary to (not in contradiction with) systems such as Givon (1983)’s, Ariel (1990)’s and Gundel et al. (1993)’s. Note that in this system, antecedent type is largely described in terms of locality, with the structural position of the antecedent (“information saliency”) and “general implicature constraints” (semantics) included as overriding factors. Since saliency is defined as a factor that can (but doesn’t have to) override locality constraints, there is nothing in Huang (1994)’s system that is incompatible with Kaiser & Trueswell (2008)’s Form-Specific Multiple Constraints Hypothesis either. It may very well be the case that the GCIAs for different anaphors are overridden more often by one factor (structural position) rather than another (semantic factors); Huang (1994)’s system is neutral to such claims.

Huang (1994) provides ample examples that this analysis explains patterns of anaphora in Mandarin Chinese, and it is not difficult to extend the analysis to other null argument languages,
including some Romance languages. Under this GCI analysis, overt pronouns will trigger a sort of “default” topic-shift (or anti-local) interpretation in such languages. This is not to say that overt pronouns in GCI-type null-argument languages always result in topic-shift interpretation. Crucially, as conversational implicatures, GCIs are cancelable, and Huang (1994)’s system includes the Consistency Constraints as a mechanism by which the default topic-shift implicature may be superseded by other factors, including general discourse context.

### 3.6 Summary

This chapter has explored in further detail the claims about topic-shift and overt pronouns made in the Interface Hypothesis studies. We have seen that the association of topic-shift with overt pronouns in Romance(-like) languages has been made in both empirical and theoretical work. We explored anaphora scales as a particularly useful way to look at the overt/null distinction, and then made a connection between these scales and Gricean analyses of anaphora. While these anaphora scales claim universality in their underlying mechanisms, they also allow for variation, and note (but leave partially unresolved) particular variation among languages for referring expressions on the higher end of the scale (i.e., various pronominal forms, as opposed to constructions with lexical NPs). The Form-Specific Multiple Constraints Hypothesis also suggests that different anaphors may be sensitive to different kind of antecedent properties; Filiaci et al. (2013) extend this analysis to explain differences in overt pronoun interpretation in Italian and Spanish.

---

16 A note on Chinese: Even though Chinese and Japanese are often grouped together (since both languages have object drop and lack agreement morphology), null forms are more common in the latter than the former. Gundel et al. (1993), for example, found that 70% of in-focus (=current topic) referents were coded with nulls in Japanese, compared to only 28% in Mandarin and 19% in Spanish. Furthermore, a study by Jia & Bayley (2002), comparable in methodology to the variationist studies cited above, found that overt pronouns favored switch-reference with a weight of .56, similar to Spanish.
Under these analyses, the task of the L2 learner is to acquire the anaphora scale for their L2, a task which is complicated by the variation among anaphors on the higher end of the scale, a problem which is still being investigated. The FSMC approach suggests that different constraints on antecedent saliency may play a role in distinguishing anaphors that otherwise seem to overlap; Filiaci et al. (2013) extend this idea to differences among similar anaphors in two different languages. What L2 learners should not have to do is learn the principles that underscore the scale in the first place; we should not see L2 learners using nulls for switch-reference more often than overt pronouns. Indeed, this is exactly what is found in the IH studies cited previously: L2 participants show more overlap in their interpretation preferences for null than overt, but do not interpret nulls for switch-reference any more frequently than L1 participants (and certainly not more often than overts).

The next question is how and whether the theories of anaphora explored above can be useful in explaining the ambivalent interpretation of overt pronouns in Japanese found in various experimental studies. In the next chapter, I explore the empirical data on Japanese 3rd person pronoun use in more detail, describe the properties of overt pronouns in Japanese, and suggest that the distinction between generalized and particularized conversational implicatures may be useful for modeling the difference between overt pronoun use in Japanese and in other languages.
CHAPTER 4

OVERT PRONOUNS IN JAPANESE

4.1 Topic-shift Effect in Japanese: Does It Exist?

Japanese, like the Romance-like languages described in the previous chapters, also allows null subjects, so the question arises as to whether or not Japanese speakers also interpret overt pronouns with topic-shift/switch-reference. As briefly reviewed in Chapter 2, empirical evidence suggests that the answer is no. This is somewhat unsurprising, as there are several key differences between Japanese and these other languages. For example, as mentioned in Chapter 1, scholars have long separated “null subject” languages into two different groups: Romance-like “pro-drop” languages (null subjects and agreement morphology) and East Asian “zero-topic” languages (null subjects and objects17, no agreement), with Japanese in the latter category (Jaeggli & Safir, 1989; Huang, 1984; 1989). While these studies focused on the properties and distribution of null forms, it is also important to note that the properties and distribution of overt pronouns vary among these languages as well. For example, Japanese overt pronouns are very infrequent; the rate of nulls is much higher to begin with, and overt reference is much more commonly made with a lexical noun, including titles and demonstrative phrases such as ano hito (“that person”), even when referring to first- or second-person (although I focus on 3rd person pronouns in this study). In this chapter, I first review the experimental evidence suggesting that Japanese overt 3rd person pronouns do not consistently elicit topic-shift to the same extent found in other null argument languages, before exploring the general properties of Japanese pronouns, and then suggesting that the Gricean distinction between particularized and generalized

17 It is not often mentioned in these studies, but overt pronouns in Japanese are also seldom found in non-argument positions, such as possessive pronouns, where they are commonly found in English.
conversational implicatures can best capture the distinction between Japanese and other null argument languages in terms of overt 3rd person pronoun use.

4.2 Previous Studies on Japanese: Little Evidence of Topic-shift

4.2.1 L1 Control Data in SLA Studies

As mentioned in Chapter 2, Okuma (2012), using the standard Interface Hypothesis experimental procedure, found inconsistent topic-shift for both native and non-native participants (Okuma, 2012, p. 7):

(11) *Okaasan-wa musume-ni kanojo-ga/o koooto-o kiru toki-ni kisu-o sita.*

mother-TOP daughter-DAT she-NOM/o coat-ACC put.on when kiss-ACC did

“The mother kissed the daughter when she/o put on the coat.”

Whereas Sorace & Filiaci (2006) found that L1 Italian participants choose a subject interpretation for overt pronouns in such contexts only 8% of the time, L1 Japanese participants chose a subject interpretation 40% of the time.\(^{18}\) In other words, Italian speakers shifted the topic from the subject to another referent 92% of the time, whereas Japanese speakers did so only 60% of the time; for comparison, Sorace & Filiaci (2006)’s L2 participants chose topic-shift interpretations 79% of the time, more often than the L1 Japanese speakers. (Note that the two language groups also differ in the Null condition, with the subject-antecedent bias stronger in the Japanese data).

\(^{18}\) For this table, I chose the conditions from each study that were most comparable to each other (forward anaphora, no scrambling, no topic-marking on non-subject referents).
Table 7. Results of Sorace & Filiaci, 2006 and Okuma (2012), Compared

<table>
<thead>
<tr>
<th></th>
<th>Overt</th>
<th></th>
<th>Null</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Subject</td>
<td>Complement</td>
<td>Other</td>
</tr>
<tr>
<td>Sorace &amp; Filiaci (2006)</td>
<td>8%</td>
<td>82%</td>
<td>11%</td>
</tr>
<tr>
<td>Okuma (2012)</td>
<td>40%</td>
<td>45%</td>
<td>15%</td>
</tr>
</tbody>
</table>

While neither the Italian group nor the Japanese group shows an English-like interpretation for overts (i.e., a trend towards subject-interpretation), the Japanese group does not show the same overwhelming non-subject interpretation tendency as the Italian group. Finding no statistically significant difference in the overt condition, Okuma (2012) ultimately concludes that there is no antecedent preference for overt subject pronouns in Japanese. In their choice of a same-subject vs. non-subject referent, the L1-Japanese participants in Okuma (2012)’s study are effectively at chance. It is worth noting that Okuma (2012) also contained additional conditions testing the effect of antecedent topic-hood and surface word order on overt pronominal interpretation, and found no effect, with L1 participants at chance across the board.

Kanno (1997)’s study is particularly interesting, since her items lacked context completely. Recall from Chapter 2 that participants were given a single sentence with an embedded clause, and asked whether the referent of the embedded pronoun was the matrix subject or “someone else,” with the option to choose both.

(12) Tanaka-san-wa raishuu kare-ga Toukyou-e iku-to itteimashita yo.
Tanaka-Mr.-TOP next.week he-NOM Tokyo-to go-that said SFP
“Mr. Tanaka said that he would go to Tokyo next week.”

Results for L1 Japanese participants showed no clear preference for either a same-subject or a topic-shift (“someone else”) interpretation, with participants choosing a same-subject or “either (a) or (b)” interpretation 47% of the time, and the “someone else” option 53% of the time. As in Okuma (2012)’s study, participants were more or less at chance.
As described in Chapter 3, Carminati (2002)’s Position-of-Antecedent Hypothesis claims that in Italian, speakers look to subject position (Spec-IP) for the antecedent of a null pronoun, and other positions for the antecedent of an overt pronoun. While her experimental data largely supported the PAH, Carminati (2002) noted that the effect is strongest when overt pronouns appear in ambiguous contexts, and seems to disappear in unambiguous ones (Carminati, 2002, p. 303). This does not seem to be the case in Japanese, as participants in both Okuma (2012) and Kanno (1997) react to overt pronouns in ambiguous contexts with an ambiguous interpretation pattern.

4.2.2 Evidence from Native Speaker Production Studies

An L1 production study by Ueno & Kehler (2010) provides further evidence that Japanese overt 3rd person pronouns have an ambiguous interpretation status, as their study also showed no preference for a topic-shift when Japanese native speakers were given an overt pronoun prompt. Items included a contextual sentence followed by a null subject prompt, a 3rd person pronoun prompt, or a “free” prompt in which participants were directed to choose their own subject (Ueno & Kehler 2010, p. 2058):

(13) Context: *Taro-wa Jiro-ni hon-o watashita tokoro datta.*  
     Taro-TOP Jiro-to book-ACC handed scene was  
     “Taro handed Jiro a book.”

Prompt: ø/Kare-wa/Free__________________________________.

The experiment included two additional conditions: topic vs. subject marking on the first NP (*Taro* in the example above) and imperfective vs. perfective aspect on the verb (*watashita* in the example above). Collapsed across aspect, participants were more or less 50-50 in their choice to continue the sentence in the prompt with *kare* referring to *Taro* or *Jiro*, regardless of whether the NP is marked subject or topic. With the aspect condition factored in, participants tended to
choose *Jiro* (i.e., topic-shift) more often in the perfect at a rate of 55%, and tended to choose *Taro* (i.e., no topic-shift) more often in the imperfective, at a rate of around 60%. The authors conclude that overt pronouns do not display a non-subject bias; again, they are more or less at chance. Ueno & Kehler (2010) use these results to argue against the application of the Givenness Hierarchy to Japanese 3rd person pronouns, noting that the presence of an overt pronoun does not implicate that the interpretation should be different than that of a null.

Furthermore, Ueno & Kehler note a crucial difference between null and overt pronouns, in that the latter seem more constrained by pragmatic factors than the former, as evidenced by the fact that subject-interpretation rates for nulls were consistently high (75-80%), whereas subject-interpretation rates for overts seemed to fluctuate at the same rate as those found in the “free” prompts (which showed 45% (imperfect) and 30% (perfect) same-subject interpretation in the aspect condition, comparable to, but lower than, 60% (imperfect) and 45% (perfect) in the overts). A follow-up study also showed a greater tendency for overts to be associated with a topic-shift interpretation when there was no implicit causality built into the context sentence, that is, more topic-shift when the predicate *surprise* (implicit causality) is replaced with *praise* (no implicit causality) in the example test item below (Ueno & Kehler, 2011):

(14) Context: *Taro-wa Jiro-o odorokashita tokoro datta.*
    Taro- TOP  Jiro-ACC surprised scene was
    “Taro surprised Jiro.”

Prompt: Ø/Kare-wa/Free ____________________________.

Taken together, these results suggest that (i) there is difficulty integrating Japanese 3rd person pronouns into an anaphora scale like the Givenness Hierarchy, and (ii) the specific context of the overt pronoun is crucial to its interpretation, in contrast to nulls, where no such influence is found. I will return to the importance of context in Section 4.4 below.
4.2.3 Evidence from Corpus Data

In a previous corpus study, I examined the use of 3rd person pronouns by native speakers of Japanese using transcripts of 15 minute interviews (Uemura, 1999; Nagano, 2012). In the seven native-speaker transcripts that I studied in detail, there were a total of nine 3rd person pronouns; one was morphologically marked as subject, four were marked as topic, and four marked as obliques. These 3rd person pronouns were used for a variety of functions, including emphasis/implicit contrast, such as in the example below, in which the speaker contrasts her homestay son with a more stereotypical exchange student, who is not expected to be of Japanese descent. Note that reference to the homestay son is maintained from the interviewer’s question, i.e., there is no topic-shift.

(15) A: [asks how homestay student is different from Japanese people]

B: sou desu-ne, ano, kare-wa nikkeijin desu-node, ano so be um he-TOP Japanese.descent be-because um mita-tokoro-wa mattaku nihonjin-na-n desu-ne. saw-place- TOP exactly Japanese-be-NOMI be “Right. Um, he’s of Japanese descent, so, um, when (you) look (at him), (he)’s just like a Japanese person.”

The corpus data also showed that 3rd person pronouns were used as a sort of fall-back when an overt form was needed in order to bear an oblique case marker, but all other overt forms had been exhausted or were otherwise unretrievable, as in this excerpt, in which the speaker describes the same referent (Maya Angelou) using three different long, modified noun phrases, before using kanojo (“she”) to get on with her utterance, and to support oblique references throughout. Note that kanojo is never marked as an argument or topic:

(16) ano, ano, kokujin-no kokujin-josei-no, ano, shijin-de um um black-GEN black-woman-GEN um poet-and maya-anjero-san-tteiu-kata, ano, kurinton-no, ano, daitouryou-no Maya.Angelou-Ms.-called-person um Clinton-GEN um president-GEN shokuninshiki-de chi-o yonda-kata na-n desu-keredomo, kanojo-ni
inauguration-at poetry-ACC read-person be-NOMI be-but she-in
totemo kyoumi-o motteite, ima kanojo-no hon-ni deatte, sugoku,
very interest-ACC hold.and now she-GEN book-to encounter very
maa, kanojo-nituite yatteikita-naa-tteiu-fiu-ni sugoku omotte, de, maa,
well she-about do.want-hmm-called-way-in very think.and and well
iroiro-na kakudo-kara, apuroochi shite, maa, mochiron bungaku
various corner-from approach do.and well of.course literary
sakuhin-o mitari-toka kanojo-no kaku, maa, bungaku sakuhin-o
works-ACC see-etc. she-GEN write well literary works-GEN
mitari-da-toka...
ssee-be-etc.
“Um, um, (it)’s on a female black poet, a lady called Maya Angelou, the lady who
read poetry at President Clinton’s inauguration, but (I)’m very interested in her, now
(I)’m reading her books, and well, (I)’m thinking (I)’d like to do (my thesis) on
something about her, and, well, there are various angles, approaches to take, well, of
course (I)’ll look at (her) literary works, the literary works that she wrote and…”

Finally, there was one instance of the pluralized form kare-ra “they” that seems to be used for
topic-shift, but in a complex, utterance-long fashion that also mixes in an implicit contrast. The
speaker here contrasts ippan-no-hitobito (“average people”), thereafter referred to as jibun-tachi
(“selves”), with the members of the Aum cult, who she refers to consistently as kare-ra (“they”).
In this case, all three instances of kare-ra are marked as topic (two with topic-marker –wa, one
with the topicalizing construction da-tte):

(17) ippan-no-hitobito-ga oomu-o dou miru-ka-toka oomu-no-hitotachi-o
average-GEN-people-NOM Aum-NOM how see-Q-etc. Aum-GEN-people-ACC
dou miru-ka-tteiu-koto na-n desu-kedo, nanka kou jibun-tachi-wa
how see-Q-called-thing be-NOMI-be-but like like this selves-TOP
jibun-teiou-no-va sono-ippan-no-hitotachi desu-ne jibun-tachi-wa
self-called-NOMI-TOP that-average-GEN-people be selves-TOP
seijou-de karera-wa kyoujin-de karera-wa yowai-kara, shuukyou-ni-wa
normal they-TOP crazy-and they-TOP weak-because religion-in-TOP
matta— jibun-tachi-wa sou janai zettai daijoubu-tteiu souiu nanka kou
exa— selves-TOP so be.NEG absolutely fine-called that like this
wakete kangaeru kangekata-ga sugoku masukomi-toka-de-mo tonikaku
divide.and think think.way-NOM very media-etc.-on-too anyhow
oomu-o hinan shita, hihan shita-tteiu kanji-ga shite, watashi-wa
Aum-ACC criticize do judge do-called feeling-NOM do.and I-TOP
sore-wa doumo, nattoku-ga ikanai-tteiu-ka jibun futsuu-no-hi—
that-TOP somehow assent-NOM go.NEG-called self regular-GEN-PE—
goku futsuu-no, karera da-tte motomoto goku futsuu-no hitotachi-de…
very regular-GEN they be-that originally very regular-GEN people-and “(That)’s how average people; view Aum, how they view the Aum people, but… (it)’s like this, (them)selvesi, (them)selvesi meaning those average people, (them)selvesi are normal and theyj are crazy, and theyj are weak, so in religion definit—(them)selvesi aren’t like that, (they)’re definitely okay, that kind of, like, way of thinking where (it)’s neatly divided, (it)’s very— the media, too, there’s a feeling of criticizing, of judging Aum no matter what, and I’m not really satisfied with that, (I) guess, (them)selvesi/(my)selfj normal people, very normal, theyj were originally very normal people…”

In sum, the corpus data showed that 3rd person pronouns were (i) rare, and (ii) used for various functions including, but not limited to, topic-shift, with no overwhelming trend for particular interpretation pattern (there are barely enough 3rd person pronouns in the data to compute a statistical trend to begin with—I exhaust all but one of the tokens in the excerpts above). This corpus data fits in with the experimental studies showing mixed interpretations of Japanese overt 3rd person pronouns.

In sum, experimental and corpus studies reveal that Japanese overt 3rd person pronouns are rare in natural conversation data, and that, unlike null pronouns, they do not have a consistent antecedent preference, with participants choosing subject and non-subject antecedents at chance in experimental studies (Kanno, 1997; Okuma, 2012; Ueno & Kehler, 2010). Ueno & Kehler (2010) argue that the Givenness Hierarchy does not fit their results for Japanese overt pronouns, since they do not implicate the opposite interpretation of nulls in their experiment. Furthermore, the fact that Japanese overt pronoun interpretation remains at chance in ambiguous contexts contrasts with results for Italian in Carminati (2002)’s PAH studies, which showed overt pronoun interpretation conformed to the PAH (i.e., preferred a non-subject antecedent) more often in ambiguous contexts than in unambiguous contexts. That is, in Italian, the GCI for overts clearly emerges in ambiguous contexts but can be overridden when additional context and/or other semantic information is provided, but in Japanese no default bias emerges in ambiguous contexts in the first place. This argues against a generalized conversational implicature (GCI) analysis
along the lines of Huang (1994) for Japanese overt 3rd person pronouns, since Japanese overt pronoun interpretation does not seem to reflect any sort of default bias. Furthermore, unlike in the Italian studies, context does not seem to override any interpretation bias, but instead helps to build up an interpretation in the first place, as found in Ueno & Kehler (2010; 2011). Below, I explore the properties of overt pronouns in Japanese, with a special emphasis on 3rd person pronouns, before arguing that a particularized rather than generalized conversational implicature analysis best explains the interpretation of overt 3rd person pronouns in Japanese.

4.3 Properties of Overt Pronouns in Japanese

4.3.1 Japanese Pronouns as Nouns

Many Japanese grammarians have questioned the need for a morphosyntactic category of pronoun separate from that of noun (Kuroda, 1965; Noguchi, 1997). Kuroda (1965) describes four properties of Japanese pronouns that most diverge from their English counterparts, claiming that while the former may share certain semantic features with the latter, morphosyntactically they are different creatures. Specifically, Kuroda (1965) cites (i) lack of special inflection paradigms for case and number, (ii) direct modification by adjectives, (iii) non-unique forms for first and second person, (iv) different sociopragmatic “connotations” for different pronouns. Ultimately, Kuroda argues that the Japanese equivalent of English (overt) pronouns are null, unpronounced forms, and that what are considered to be (overt) personal pronouns in Japanese are lexical items within the category of noun that happen to share certain semantic features with English pronouns (Kuroda, 1965, p. 108, 123). There have been more recent analyses along these lines, such as Noguchi (1997)’s analysis that Japanese pronouns are NPs that bear a pronominal feature, while their English counterparts are DPs with a pronominal feature, but the general
theme has remained the same since the early 20th century: there is little reason to posit pronoun and noun as separate morphosyntactic categories in Japanese (see Obana, 2003 for an overview).

Indeed, unlike English pronouns, which retain case and number paradigms where regular nouns do not (i.e., I, me, my, mine, we, us, our, ours), Japanese pronouns inflect for case and number no differently than other nouns, by directly attaching case morphemes such as nominative –ga or accusative –o, and, for animate NPs only, the (largely interchangeable) plural suffixes –tachi and –ra:

(18a) ore-ga omae-o mita
    I-NOM you-ACC see.PAST
    “I saw you.”

(18b) kodomo-ga gakusei-o mita
    child-NOM student-ACC see.PAST
    “The child saw the student.”

(18a’) ore-tachi-ga omae-ra-o mita
    I-PL-NOM you-PL-ACC see.PAST
    “We saw you guys.”

(18b’) kodomo-tachi-ga gakusei-ra-o mita
    child-PL-NOM student-PL-ACC see.PAST
    “The children saw the students.”

Kuroda (1965) gives the following example of adjectival modification, of a type that is not allowed in English (p. 105):

(19) chiisai kare
    small he
    “the small he”

It is also worth noting that pronouns can be modified using demonstratives and relative clauses, such as:

---

19 Note that in the absence of plural marking, Japanese NPs are ambiguous as to plurality; (18b) could also be glosses as “Children saw students.”
In other words, Japanese pronouns can be modified in the same way as nouns.20

Furthermore, there are multiple options for pronouns in the first- and second-person, most of which are etymologically tied to nouns, such as 1st person pronoun boku, meaning “servant,” and 2nd person kimi, meaning “prince” (Martin, 1975, p. 1075-1080). Below is a partial list of the pronouns discussed in Martin (1975):

<table>
<thead>
<tr>
<th>First-person</th>
<th>Second-person</th>
</tr>
</thead>
<tbody>
<tr>
<td>watakushi, watashi, atakushi, atashi, washi, ore, boku, jibun, uchi</td>
<td>anata, anta, omae, kimi, temae, otaku, onushi</td>
</tr>
</tbody>
</table>

The differences among these various forms are sociopragmatic, not morphosyntactic. Gender, social status, and familiarity all play a role in pronoun selection, with some variations in regional and other social dialects (as well as historical factors, as certain pronouns may be associated with characters in period dramas or pre-modern literature). The 2nd person pronoun omae (a nominal form meaning “in front of”), for example, is considered intimate with friends or younger family members, but inappropriate and even condescending for use with a social superior or stranger; some 2nd person forms, such as the originally-respectful temae (a nominal form meaning “to the side of”), are now even considered vulgar. Crucially, gender factors in 1st and 2nd pronoun

---

20 While English pronouns occasionally allow adjectival modification (stupid me) and relative clause attachment (You, who built this nation with your hard work and effort, deserve these accolades), these types of modification are more restricted, in terms of morphology, productivity, and/or formality. For example, note that adjective-modified pronouns must be in oblique case (*Stupid I sent him an e-mail after swearing never to speak to him again).
selection are not morphologically constrained; a woman using the traditionally masculine pronoun *boku* is viewed as making a deliberate social choice, not a grammatical “error”; *watakushi* is used by men in more formal registers, and women in both formal and casual registers. Additionally, the person features of 1<sup>st</sup> and 2<sup>nd</sup> person pronouns may be variable; that is, the 1<sup>st</sup> person pronoun *boku* may sometimes be used as a form of address for a second person (usually a young boy); the pronoun *temae* has historically been used for both 1<sup>st</sup> and 2<sup>nd</sup> person reference (Martin, 1975). In sum, the descriptive data support an analysis that treats Japanese overt pronouns as a special subclass of nouns that happen to bear some semantic features similar to those of pronouns found in languages like English. For the remainder of this chapter, I focus on 3<sup>rd</sup> person pronouns, the main focus of this study.

### 4.3.2 Third-person Pronouns and Demonstratives

Unlike the numerous, nominal-derived first- and second-person forms, there are only two 3<sup>rd</sup> person pronouns in Japanese, derived from demonstratives, with a built-in gender distinction. While these characteristics make 3<sup>rd</sup> person pronouns sound similar to their English counterparts, this is by design. While there have always been forms used for first- and second-person reference (see Martin, 1975 for discussion of proto-forms *ba* and *na* for 1<sup>st</sup> and 2<sup>nd</sup> person pronouns), the 3<sup>rd</sup> person forms *kare* and *kanojo* were specifically adopted in the 19<sup>th</sup> century for the purpose of translating *he* and *she* in Western literature.\(^{21}\) *Kare* has its roots in an archaic demonstrative, equivalent to the modern Japanese *a*- (distal) demonstrative series, and was originally used to refer to high-status members of both sexes; *kanojo* was invented by analogy with this form.

---

\(^{21}\) Despite the intentional parallels of *kare* and *kanojo* to Western 3<sup>rd</sup> person pronouns like *he* and *she*, Japanese has no 3<sup>rd</sup> person form equivalent to *it*, with the *s*-series demonstrative *sore* (“that-proximal thing”) sometimes used as a stand-in for word-by-word translation.
As such, much of the discussion of these 3rd person forms, especially in the Japanese-language literature, has analyzed them as demonstratives, or has sought to place them somewhere in the Japanese tripartite demonstrative system (Okamura, 1972; Kinsui, 1989, among others). For example, Takubo & Kimura (1992), working under this tradition, argue that each of the three demonstrative series corresponds to a different information status for the referent: the k-series (“this”) is used when the entity is known to the speaker only; the s-series (“that-proximal”) is used when the entity is known to the hearer only; the a-series (“that-distal”) is used when the entity is known to both. Since, according to their analysis, overt 3rd person pronouns kare and kanojo can only refer to (i) an entity known to both the speaker and hearer prior to the beginning of the conversation, or (ii) an entity that is physically present in the environment during the conversation, they operate similarly to the a-series (Takubo & Kimura, 1992, p. 137). Additionally, Hoji (1991) draws on the tradition of analyzing kare as a demonstrative in his analysis of the Overt Pronoun Constraint (OPC) in Japanese, arguing that kare’s resistance to being bound under a quantifier is a direct result of its lingering deictic nature.

However kare and kanojo differ from modern demonstratives on several points. First of all, other demonstratives can be used for person reference while maintaining a more general demonstrative meaning. For example, the series kochira (“over here”), sochira (“over there-proximal”), and achira (“over there-distal”) can be used for 1st, 2nd, and 3rd person reference respectively, but this is a specialized use of their more generalized meanings. Kare and kanojo, on the other hand, cannot be used as more general demonstratives. Even more specialized demonstratives, such as the colloquial forms koitsu (“this guy/gal”), soitsu (“that-proximal

---

22 In fact, since kanojo was originally introduced as a written form (彼女), its pronunciation varied between ka-no-onna and ka-no-jo, depending on which of two possible readings was chosen for the second character, before it stabilized as the latter.
guy/gal”), and aitsu (“that guy/gal-distal”) differ morphosyntactically from kare and kanojo. In particular, the former cannot be directly modified by a demonstrative adjective, while the latter (like 1st and 2nd person pronouns) can, such as the form sono kare (“that he’’). Furthermore, kare and kanojo have been adopted as lexical NPs in their own right, meaning “boyfriend” and “girlfriend” without any morphosyntactic adjustment; no demonstrative forms have become lexicalized in this way. Furthermore, Obana (2003) argues against the intuitions in Kimura & Takubo (1992), claiming that unlike an a-demonstrative, kare/kanojo (i) cannot be used for an unknown referent salient in the immediate environment, and (ii) that the referent of kare/kanojo need only be known to the speaker (p. 141-142). Obana ultimately concludes that kare/kanojo cannot be easily classified along with other determiners, and that their grammatical status remains ambiguous, a point that I will emphasize in Section 4.4.

4.3.3 Sociopragmatic Constraints on Third-Person Pronouns

Japanese 3rd person pronouns are, like their first- and second-person counterparts, also constrained by sociopragmatic factors. Hinds (1975) explored sociopragmatic constraints on kare/kanojo, surveying Japanese young adults and teenagers on their use of 3rd person pronouns. From these surveys, Hinds (1975) made the following observations: (i) young adults used 3rd person pronouns more often than teenagers, reflecting an association with education and worldliness, (ii) women used 3rd person pronouns more than men, (iii) 3rd person pronouns were not used to refer to family members, social superiors, or public figures, (iv) 3rd person pronouns were found more often in translations of Western languages than in everyday conversation, and

---

23 This use of kare and kanojo creates some problems for empirical analyses. First of all, it can muddle judgments of direct modification (chiisai kare, sono kare), since out of context these can always be interpreted as the lexical NP forms (“small boyfriend”; “that boyfriend”). Second, it can cloud corpus data on 3rd person pronoun usage rates if the lexical uses are not sorted out.
(v) using 3rd person pronouns too often was thought to be “improper.” He also noted the use of *kare/kanojo* as lexical nouns meaning “boyfriend” and “girlfriend,” which he considered “emotional connotations” (p.155). Overall, Hinds (1975) concluded that 3rd person pronouns were marked forms and that speakers had conscious opinions about their use.

Obana (2003) re-examined the use of *kare/kanojo* a generation later in order to investigate if and how attitudes towards 3rd person pronouns had changed since Hinds’ original 1975 study. The results showed a striking effect of age: participants under the age of 45 were much more accepting of 3rd person pronouns that those 45 and older (i.e., the peers of the participants in Hinds’ original study). Additionally, Obana (2003) found three main changes in use since Hinds (1975): (i) younger participants are more comfortable using 3rd person pronouns to refer to family members (but not to parents), (ii) there is no longer any gender difference in terms of who uses 3rd person pronouns more often, (iii) there is now more acceptance of using *kare/kanojo* to refer to public figures. Obana (2003) also concludes that overall there is now less “psychological effect” of using *kare/kanojo* than in 1975, and that the words have become more “objective.” In the same study, Obana followed up the survey data with twenty-five interviews examining 3rd person pronoun use in actual conversation. Results showed (i) *kare/kanojo* were used sparsely in comparison to null forms, lexical nouns, and demonstratives, (ii) the greatest number of interviewees used *kare/kanojo* when referring to a friend/colleague, followed by celebrities, with only one interviewee using them to refer to a family member. She also noted a tendency to use 3rd person pronouns to indicate a shift in viewpoint toward the referent, such as introducing an objective judgment about that person. Overall, Obana (2003) notes a trend towards 3rd person pronouns becoming more “neutral” in connotation and less marked in use
over the 25 years between her study and Hinds (1975). These studies taken together suggest that
the use of the relatively young *kare* and *kanojo* is still evolving into the present.

4.3.4 Infrequency of Overt Pronouns

As noted above, Kuroda (1965) claimed that null forms in Japanese are the
morphosyntactic equivalent of overt pronouns in English. This fact is reflected in the infrequency
of overt pronouns in the former language compared to the latter. In the corpus study described in
Section 4.2, I found at total of .5% of predicates accompanied by a 3rd person pronoun in the
seven transcripts studied in detail; there were only 21 3rd person pronouns in all forty fifteen-
minute interviews in the corpus (Uemura, 1999; Nagano, 2012). First-person pronouns, which
serve discourse organizational functions such as floor-taking, were somewhat more frequent (a
rate of roughly 4 1st person pronouns per 100 predicates in the Uemura corpus), while 2nd person
pronouns were particularly rare (zero uses by interviewees in the entire corpus), as they are
usually shunned for sociopragmatic reasons in favor of a name, title, demonstrative, or noun.
Another corpus study by Amano & Kanno (2000) (in Ueno & Kehler 2010) found 43,004 3P
pronouns in 11,073,167 newspaper sentences, a rate of .004 pronouns per sentence, a rate
comparable to what I found in spoken data.

Hinds (1975), in addition to questionnaire data, elicited spoken data using blank cartoons
with a male and female character (a context he claimed would force English speakers to use *he*
and *she*), and found only 18% of participants used *kare* or *kanojo* at least once. Instead,
participants used the terms *okusan* “wife” and *goshujin* “husband” to refer to the characters, or
simply made up names for them. Obana (2003)’s interviews were designed to elicit 3rd person
reference from participants (with questions about family members, friends, celebrities, etc.), and
she did indeed find that 19 out of 25 participants used *kare* or *kanojo* at least once. However, she
notes that the total number of $3^{rd}$ person pronouns was still low, with participants preferring to use a name, demonstrative, or null form for $3^{rd}$ person reference. Recall also from Chapter 3 that the rate of Japanese null pronoun use is much higher than in other null-argument languages, including Spanish (Romance-like) and Chinese (long assumed to be similar to Japanese).

4.3.5 Summary: Third-Person Pronouns as Infrequent, Indeterminate Forms

Third-person overt pronouns in Japanese are much rarer than their English counterparts (and rarer than their counterparts even in other null argument languages). In addition to their great infrequency, Japanese $3^{rd}$ person overt pronouns display mixed lexical properties, bearing the semantic features of pronouns, conforming to the morphosyntactic properties of lexical nouns, and retaining traces of their relatively-recent etymological roots as demonstratives in their interpretation and use. In the next section, I argue that the relative scarcity of $3^{rd}$ person pronouns invites the same conversational implicature approach to their use found in other null-argument languages, but that their mixed lexical properties suggest a particularized, not generalized, conversational implicature analysis.

4.4 Capturing the Difference between the Two Language Types: A Neo-Gricean Approach

4.4.1 A Particularized Conversational Implicature (PCI) Approach for Japanese

In Chapter 3, I described Huang (1994)’s Neo-Gricean account of anaphora in Chinese. Huang (1994), following Levinson (1987; 2000), describes a generalized conversational implicature (GCI) analysis of anaphora, in which various anaphors (null pronouns, overt pronouns, demonstratives, full NPs) implicate local or non-local coreference based on principles derived from Gricean Maxims. Under the GCI approach, all other things being equal, the overt pronoun should trigger a topic-shift interpretation (M-implicate the complement of the same-subject zero-anaphora interpretation), whenever an overt pronoun is used where a null form
could be used instead, although this implicature can be cancelled by various factors, including antecedent saliency, world knowledge, and other contextual information. This is exactly what we find in the Interface Hypothesis monolingual control data. Carminati (2002) specifically notes that the topic-shift bias predicted for overt pronouns by the PAH seems to be strongest in ambiguous contexts, not those in which contextual or semantic factors (such as gender) can provide a clearer interpretation. We would expect this to happen under the GCI approach: when there is enough contextual information to override the implicature, the implicature disappears.

The empirical data for Japanese found in Kanno (1997), Okuma (2012), and Ueno & Kehler (2010), however, do not fit well with the GCI approach, since they show no evidence of such a clear-cut interpretation preference for Japanese overt pronouns, even in ambiguous sentences.

I argue that the GCI analysis is a poor fit for Japanese overt 3rd person pronouns because Japanese overt 3rd person pronouns do not fit well on an anaphora scale, which lies at the core of the GCI approach (recall that the M-implicatures in Huang (1994)’s system are based on “stronger” forms, such as overt pronouns, in competition with “weaker” forms, such as nulls; this ranking of anaphors is precisely what is found on various anaphora scales). As described in section 4.3 above, overt 3rd person pronouns in Japanese display mixed lexical properties, with the semantic features of pronouns, morphosyntactic features of nouns, and traces of their etymological origin as demonstratives. Pronouns, lexical nouns, and demonstratives normally form separate points on an anaphora scale, such as those underlying GCI analyses of anaphora; the fact that Japanese 3rd person pronouns demonstrate properties of all three suggest that they do not fit easily onto such a scale, and therefore a GCI account will not suffice.

Indeed, looking at the criteria offered for the ranking of anaphors on the various scales in Chapter 3, it is difficult to find a place for Japanese 3rd person pronouns. In terms of
phonological weight, they are quite heavy (kanojo is three morae; kare is two); Japanese functional morphemes are usually one mora (such as case particles (i.e., ga-subject marker), complementizers (to, “that”), and postpositions (ni “to”; de “by, at”), with a handful of two-mora postpositions (kara “from”, made “until”). Under the phonological content criterion, therefore, they are similar to lexical nouns. In terms of semantic content, they are similar to pronouns found in other languages in terms of gender and person features, but may also retain traces of their demonstrative origins, in addition to their associated lexical meanings “boyfriend” and “girlfriend.” In terms of phonological content, therefore, Japanese 3rd person pronouns are heavy, comparable even to nouns. In terms of semantic weight, they are more blanched than nouns, but not quite as blanched as the typical 3rd person pronoun in English.

These mixed lexical properties, taken in conjunction with the ambivalent interpretation data, suggests that Japanese overt 3rd person pronouns may trigger particularized (rather than generalized) conversational implicatures, based on the second part of the Maxim of Quantity. The rarity of overt 3rd person pronouns is such that their presence where a null is expected would trigger a conversational implicature (since speakers are saying “more than is required”); this much is the same as the GCI system. However, because the properties of Japanese 3rd person pronouns are mixed, they are not integrated into the GCI-based anaphora system in Japanese (as opposed to nulls, demonstratives, reflexives, and lexical NPs, which are). As such, there is no a priori tendency for overt Japanese pronouns to be interpreted as topic-shift (or anti-local) as are overt pronouns in languages covered by the GCI analysis. Because of their rarity, using a 3rd

24 While these forms could be considered merely homophonous, some participants in a pilot study of my experiment informally mentioned attempting to construct interpretations based on these meanings.

25 In Huang (1994)/Levinson (2000)’s system, this would be analogous to an I-implicature, but they do not discuss PCIs in detail in their work.
person pronoun where a null form is expected does create a conversational implicature, but the exact nature of this implicature must be calculated anew with each utterance based on the specific context of the utterance, with no predisposition towards a topic-shift reading over another kind of implicature. The calculated PCI could be topic-shift, but it could also be something else, such as emphasis, or even a socially-distancing effect that has only an indirect effect on reference resolution.

On the other hand, GCIs for overt pronouns have a default interpretation: the complement of the maintain-reference bias for nulls, which is topic-shift. There are, of course, non-topic-shift interpretations for overt pronouns in languages that fall under the GCI approach; crucially, however, in the GCI case, these effects result from the cancellation or overriding of the GCI, as opposed to the PCI analysis, where there is nothing to cancel in the first place. While these two possibilities may seem difficult to tease apart in the data, the difference should be clear in the test case where ambiguous (or no) context is provided, in which case GCI overts will bias a topic-shift reading, but PCI overts will not, leading to ambiguity in interpretation judgments. I have argued that this is the case by comparing results from Carminati (2002) to previous studies of Japanese overt pronouns.

Indeed, the experimental data on Japanese 3rd person pronouns reported in this chapter conform to this analysis. As mentioned above, Ueno & Kehler (2010) specifically argue against the applicability of the Givenness Hierarchy to Japanese 3rd person pronouns, and argue for the importance of pragmatic factors in their interpretation (creating the interpretations in the first place, not cancelling an implicature, as in the GCI case). Furthermore, under the PCI analysis, the lack of any context for an overt pronoun will lead to ambiguity, as is found in Kanno (1997)’s study. Unless the context consistently supports a subject-antecedent or a non-subject
antecedent, interpretations of sentences with overt pronouns will remain ambiguous. For example, the test items in Okuma (2012), in line with previous IH studies, included a third, external referent, which may have mitigated the usefulness of topic-shift as an explanation for the apparent Maxim of Quantity violation; it is not useful for the pronoun to be telling the hearer to shift from the subject when it is not clear which referent (indirect object or external referent) they should shift to. A lexical NP, for example, would be a much better choice for that purpose. It could be that without a context consistently supporting topic-shift (or consistently supporting some other implicature), participants ended up interpreting the overt pronoun at chance.

This analysis is slightly different from Kaiser & Trueswell (2008)’s FSMC approach. I am not arguing that Japanese pronouns are sensitive to different saliency factors than other anaphors in Japanese (or any other language), but that they are not sensitive to any saliency factors, since they are not on an anaphora hierarchy at all. Under the PCI approach, it is discourse context alone that creates their interpretation in the first place. From another angle, this interpretation could be seen as FSMC taken to an extreme: the interpretation constraints on overt 3\textsuperscript{rd} person pronouns in Japanese are so form-specific that saliency factors such as grammatical status or topic-hood of the antecedent become irrelevant.\footnote{This observation is also supported by the experiment in Okuma (2012), which found no effect of topic-hood or word order of the antecedent on overt pronoun interpretation.}

4.4.2 Previous Attempts to Put kare and kanojo on Anaphora Scales

Previous experimental studies on Japanese pronouns within the context of anaphora scales have been mixed. Hinds (1983) noted the rarity of Japanese pronouns and their mixed lexical properties in a corpus study that applied Givon (1983)’s Topic-continuity Hierarchy to Japanese. However, Hinds (1983) was mainly interested in showing that overt pronouns are
different from both nulls and NPs, not the nature of that difference. In line with Givon (1983)’s theory, Hinds (1983) did find that pronouns did pattern somewhere between zero anaphora and full nouns in terms of referential distance between referring expression and subject/topic (in number of clauses). However, Hinds’ pronoun category did not distinguish among different persons (1st, 2nd, and 3rd), and it is unclear whether or not it also includes demonstratives (kore, sore, are; “this,” “that-proximal,” “that-distal”). Two brief transcript excerpts are provided, but without access to more data, it is difficult to pinpoint the behavior of 3rd person pronouns specifically.27 Even more importantly, the study provides no standard deviations or other statistics reported beyond means, making it difficult to tell if there is more variation among pronouns (as would be predicted by a PCI analysis) than the other forms. Clancy (1980) presents a similar study using elicited story-telling data, but as no 3rd person pronouns were uttered, they were excluded from the analysis.

As mentioned above, Ueno & Kehler (2010) argue against the applicability of the Givenness Hierarchy to Japanese 3rd person pronouns. Gundel et al. (1993)’s data itself is also somewhat ambiguous on kare and kanojo. The authors place Japanese null pronoun under the “in focus” status with the 3rd person pronoun as “activated,” the same level as a stressed pronoun in Spanish. Their corpus data, however, shows that all instances of kare were “in focus,” meaning that use of kare patterned with that of the null in preferring the current topic of conversation as

27 Hinds (1983) also reports a strong influence of particle choice (topic, subject, or zero case marking) on the results, and without running a statistical analysis on the data it is difficult to tease the two factors apart. A subject marked pronoun, for example, has the exact same average distance as ellipses (2.7), whereas a topic-marked pronoun has an almost identical distance to a topic-marked or subject-marked NP (6.9 and 6.7 respectively). Only the non-case-marked pronoun shows a number particularly different from the others (4.9 vs. 10.8 in the non-case-marked NP).
an antecedent. It is also important to note that in all of Gundel et al. (1993)’s data, there were only four tokens of kare. In contrast, their data showed a total of 223 bare nouns, 88 tokens of zero anaphora, 34 tokens of sono N (“that-\text{prox} N”), and 9 tokens of kono N (“that N”). In line with the rarity of Japanese pronouns, although the instances of bare nouns tended to crowd towards the right end of the scale as predicted, a greater number of bare nouns were “in focus” (N = 14) or “activated” (N = 32) than the total number of 3rd person pronoun tokens (N = 4).

In sum, there is good theoretical and empirical evidence (and not much empirical evidence to the contrary) that Japanese overt pronouns do not fit well on an anaphora scale, and, as such, their interpretation is best analyzed with a particularized (not generalized) conversational implicature approach. As such, the specific context of each utterance is the main determining factor in their interpretation, even more so than for overt pronouns in other null-argument languages, in which context can override, but does not trigger, a topic-shift interpretation. Null pronouns, however, remain on the anaphora scale, and should behave in accordance with a GCI account; that is, a default interpretation for a local antecedent, unless the implicature is overridden/cancelled by one of the factors mentioned in Huang (1994)’s system (including antecedent saliency factors in addition to general context).

4.5 Experiment

The primary goal of this study is to compare the interpretation of overt pronouns by native and advanced L2 speakers of Japanese, and then to compare the results to those found in previous Interface Hypothesis studies on Italian and Spanish. In order to do so, however, we first need a basic working theory as to how overt 3rd person pronouns work in Japanese. In line with

\footnote{The authors later cite an excerpt in which kare is used for topic-shift to an “activated” referent from an “in-focus” referent. It is unclear where this data came from, and how it is separate from the previously-reported data.}
the PCI analysis offered above, I hypothesize that a context designed to support a topic-shift reading (as opposed to no or ambiguous context) will elicit a consistent topic-shift interpretation for Japanese overt pronouns by native speakers. Specifically, my test items will provide this supportive context for topic-shift interpretation in three ways: (i) providing a contextualizing introduction sentence that introduces two same-gender referents, (ii) providing images of the two potential same-sex referents on the screen, supporting their mutual relevance to the story at hand, and (iii) including test sentences with two separate clauses describing two activities that could be equally plausible if done by one person or by two separate people.

The context formulated above is designed to control for the factors in previous studies of Japanese that I believe led to ambiguous pronoun interpretations. The presence of two (and only two) same-gender referents should create a context in which topic-shift is a reasonable conversational implicature to calculate in the face of the Maxim of Quantity violation caused by the unnecessary overt 3rd person pronoun, as opposed to Kanno (1997), in which there was no established second character (just the choice “someone else”) and Okuma (2012), in which there was a third available referent. In the latter case, the hearer had little reason to believe the speaker was using the pronoun to indicate topic-shift, since it was unclear which of the two other referents to shift to; in the former, it was equally implausible that the speaker is implicating topic-shift, since there is no second referent involved in the situation. As such, while participants noticed the overt pronoun, as indicated by increased topic-shift answers compared to the null condition, they ended up choosing their responses at chance. Furthermore, ensuring that the activities in the two clauses could be plausibly done by either one or two people controls for extra contextual factors beyond the presence of the two referents (unlike Ueno & Kehler (2010)’s study, where participants were free to complete the sentence in any way they chose, therefore
giving them some control over the context of the sentence). Once an interpretation pattern is established for Japanese native speakers, we can then compare their interpretation to that of L2 speakers. I go into more detail on the research questions and methodology, including test items, in the next chapter.
CHAPTER 5

EXPERIMENTAL SET-UP: RESEARCH QUESTIONS, METHODS & PREDICTIONS

5.1 Research Questions

In this study, I investigate the interpretation of overt and null pronouns by L1 and L2 speakers of Japanese with an experiment, keeping in the PCI analysis of Japanese overt 3rd person pronouns offered in the previous chapter. As argued in that chapter, Japanese overt 3rd person pronouns are not associated with a default conversational implicature (GCI) as are their counterparts in English, Italian, and other languages. However, if a 3rd person pronoun is used where a null would suffice, it still counts as a flout of Grice’s Maxim of Quantity, and therefore must be accounted for by the hearer calculating a particularized conversational implicature (the kind of classic Gricean conversational implicature calculated from scratch at the moment of utterance). Depending on the specific context of the utterance, that implicature could be any number of things, including topic-shift.29 Under this analysis, in order to elicit a consistent interpretation from a group of speakers, the context must consistently support the same interpretation.

In this experiment, I test whether providing sentences with a certain context supporting a topic-shift implicature will result in L1 participants choosing a topic-shift interpretation for overt pronouns, where previous experiments have found ambiguous interpretation patterns. This context is designed to be a minimal context that still serves to provide the participant with enough reason to believe that the speaker is trying to indicate topic-shift by choosing to use the

---

29 Once again, note that this does not mean that GCI implicatures are not cancellable, and that GCI-style pronouns in other languages always result in topic-shift interpretations. The difference is that the GCI arises by use of the form itself, but can be cancelled by context, whereas the PCI does not arise from the form itself, and must be calculated using context.
overt pronoun, and involves two same-gender referents explicitly mentioned in an introduction sentence leading up to the sentence with the crucial Overt/Null pronoun alternation (I describe this context in more detail at the end of the previous chapter and in section 5.2.2 below). My first main research question is thus:

- Will L1 Japanese participants show a consistent (greater than chance) topic-shift interpretation for overt 3rd person pronouns in an experimental setting using a context that supports the calculation of a topic-shift implicature, or will participants remain at chance as in previous studies?

It then remains to be seen whether L2 participants will pattern with the L1 participants or not.

My second main research question is as follows:

- Given the same context, will advanced-proficiency L2 participants interpret Japanese overt 3rd person pronouns in the same way as L1 participants?

To answer these questions requires two main experimental conditions: Pronoun (Overt/Null) and Group (L1/L2), which will be described in further detail in the Methods section below.

There is one complication that adds a secondary condition. In order to investigate whether each type of Pronoun retrieves a subject or a non-subject referent, there must first be a “subject” referent available for the former interpretation. In Japanese what are referred to as “subjects” in English or other Indo-European languages can be marked with either a subject-marker (-ga) or topic-marker (-wa). Results of an informal pilot survey I conducted showed that native speakers preferred a subject-marked NP in sentences with an overt pronoun, but a topic-marked NP in sentences with a null. I therefore decided to add Particle (subject- vs. topic-marking) as a secondary condition. A third research question, therefore, emerges:
To what extent will the participants’ interpretations be mitigated by subject-marking (-ga) vs. topic-marking (-wa) on the first NP of the test sentence?

Further information on all three conditions is found in the Methods section below; predictions for each condition are found in section 5.3.

5.2 Methods

5.2.1 Overview of Experiment

In each trial, participants heard a mini-story consisting of an Introduction Sentence, a Test Sentence (featuring the Pronoun and Particle alternations), and a Concluding Sentence, followed by a Question about the Test Sentence designed to elicit their interpretation of the Pronoun. While listening to the mini-story, participants looked at a computer screen divided into four quadrants, with a different image in each quadrant, including two same-gender human figures and two inanimate objects related to the story. After hearing the Question, participants answered by clicking on one of the two human figures on the screen. Three kinds of data were recorded for each participant: Interpretation (response to the Question), Reaction Time (from offset of Question to mouse-click), and Eye-tracking (during the Test Sentence and Question). Below, I go into detail on the test items, participants, and procedure.

5.2.2 Test Items

As argued in Chapter 4, in order to elicit a strong (greater-than-chance) pattern of topic-shift interpretation by native-speakers for 3rd person pronouns, the Test Sentence must be placed in a context that supports such a reading; otherwise participants may ascribe the pronoun to another pragmatic effect, or not know what to do with it at all (leading to the ambiguity found in previous studies). This experiment provides such context linguistically via the Introduction Sentence, which sets the scene by introducing two same-gender human referents engaged in
some activity in a particular setting (riding a train, working in the office, etc.), as well as visually, as the images of the two possible referents on the screen reinforces the availability of both as responses to the Question. Below is a sample test item, along with the display that participants see on the screen.

![Image of two possible referents on a computer screen]

(1) Introduction Sentence

jon-san-to takeshi-san-wa kaigi-no-tame densha-de nagoya-e ikimasu.
John-and Takeshi-TOP conference-for train-by Nagoya-to go
“John and Takeshi go by train to Nagoya for a conference.”

(2) Test Sentence

takeshi-san-wa/ga biiru-o nondeiru-aida, kare-ga/ø keshiki-o nagametimasu.
Takeshi-TOP/NOM beer-ACC drink-while he-NOM/ø scenery-ACC view
“While Takeshi drinks a beer, he/ø views the scenery.”

(3) Concluding Sentence

inaka-no-tambo-ya hatake-wa totemo kirei desu.
countryside-‘s-rice.fields-and fields-TOP very beautiful be
“The fields and rice paddies of the countryside are very beautiful.”

(4) Question

dare-ga keshiki-o nagametimasu-ka
who-NOM scenery-ACC view-Q
“Who views the scenery?”

Figure 1. Example test item with screenshot of computer display.
Note that the Introduction Sentence does not favor either referent syntactically, as the two are conjoined in a single NP. After the Introduction Sentence, participants hear the Test Sentence, which contains the Pronoun/Particle alternations, for a total of four conditions: \textit{waNull}, \textit{waOvert}, \textit{gaNull}, \textit{gaOvert}. The Test Sentence consists of two clauses (subordinate and matrix) whose constituents follow the same general template across items:

\begin{center}
\textbf{Test Sentence Template} \\
\begin{tabular}{l}
N-\textit{wa/ga} & DO1 & V1 & CONJ & [PR-\textit{ga} or \(\phi\)] & DO2 & V2 \\
\hline
\textbf{Subordinate Clause} & & & & \textbf{Matrix Clause} & & \\
\end{tabular}
\end{center}

The first noun phrase in the subordinate clause (N) consists of a proper name referring to one of the two possible referents, which is marked with either the topic-marker \(\textit{wa}\)\textsuperscript{30} or the subject marker \(\textit{ga}\), followed by a direct object (DO1), a verb (V1), and the conjunction \textit{aida} (“while”) (CONJ). The matrix clause begins with either an overt 3\textsuperscript{rd} person pronoun (PR) or null subject (\(\phi\)), then continues with a second direct object (DO2) and verb (V2). The Overt pronoun in the matrix clause is always marked with \(\textit{ga}\), since two \textit{wa}-marked NPs in a bi-clausal sentence are always contrastive. The Concluding Sentence finishes off the story with information about the general situation and never has an animate subject. In the Question, participants are asked who did the activity in the matrix clause of the Test Sentence. In the display, participants see the two referents mentioned in the Introduction Sentence (Takeshi and John), the setting mentioned in the Introduction sentence (the train), and DO1 (beer).

Participants heard a total of 32 test items (8 in each condition), 32 items designed to collect pilot data on the Overt Pronoun Constraint, which also tests a Null/Overt pronoun

\textsuperscript{30} The two referents necessarily differ in linear order in the Introduction; this was controlled for by alternating which referent is mentioned as subject/topic in the Test Sentence (half have \textit{John}, half have \textit{Takeshi}).

\textsuperscript{31} Technically, the \textit{wa}-marked noun phrase is not part of the subordinate clause, but instead scopes over the whole sentence, as will be discussed in Section 5.3.
alternation (not reported on here) and 32 pure filler items, a subset of which resembled the test items but included a gender mismatch that disambiguated the referent of the overt pronoun. All items, filler or otherwise, included three sentences, a question, and four pictures. The test items included 16 of each gender (i.e., 16 with kare “he” and 16 with kanojo “she”), and alternated in the identity of the mentioned-referent in the Test Sentence (i.e., in the example item, this was Takeshi but in an equal number of items it was John; female items alternated between Kyoko and Susan). Items were presented in a pseudo-randomized order within four blocks which were themselves presented in a randomized order. All items were recorded by a male native speaker of Japanese; any recordings that showed evidence of contrastive intonation in the critical clause (screened using Praat software) were re-recorded.

5.2.3 Participants

L1 participants (N = 21) were recruited from a subject pool and tested at the University of Tokyo Komaba campus. A total of 22 participants were tested, but one was excluded for consistently low reaction times (mean RT below half a second). The majority of participants were undergraduate students, with a few older graduate students (mean age = 22; s.d. = 4.3 years). All participants were born in Japan and most attended school entirely in Japanese from elementary through undergraduate levels, although a few graduate students had some experience abroad as undergraduates or graduates. Most participants began studying English in middle school, in accordance with Japanese educational policy, but had no experience beyond the standard Japanese middle and high school English curricula (mean age of onset of English study = 11; s.d. = 2.6 years). The English language program in Japanese schools is largely designed around written college entrance exams, a fact reflected in participants’ self-reported proficiency ratings, which ranked Reading highest and Speaking lowest (on a scale from 1-poor to 5-
excellent), with greater variability in Speaking and Listening than Reading and Writing. Given the nature of the Japanese public school system, this level of English exposure was unavoidable, but also not far from that of the general population. Means and standard deviations for L1 self-ratings appear below:

### Table 9. Self-ratings of English Proficiency by L1 Japanese Participants

<table>
<thead>
<tr>
<th></th>
<th>Speaking</th>
<th>Listening</th>
<th>Reading</th>
<th>Writing</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Mean Self-rating</em></td>
<td>2.81</td>
<td>3.10</td>
<td>3.71</td>
<td>3.14</td>
</tr>
<tr>
<td><em>s.d.</em></td>
<td>1.05</td>
<td>1.02</td>
<td>.76</td>
<td>.71</td>
</tr>
</tbody>
</table>

L2 participants (N = 20) were mostly recruited from personal contacts and advertisements in the New York City area, with one participant recruited and tested in the same Tokyo lab as the L1 participants. All L2 participants were native English speakers with advanced proficiency in Japanese, scoring an average of 80% on a proficiency test given after the main experiment (described in Section 5.2.3 below), the items of which were culled from the three highest levels of the Japanese Language Proficiency Test, or JLPT (Japan Foundation and Japan Educational Exchanges and Services, 2012; 2012a; 2012b). Several participants reported having passed the N2 (second highest of five) and N1 (highest of five) levels of the JLPT or equivalent tests. L2 participants began their study of Japanese at an average age of 17.05 (s.d. = 4.54) years and completed an average of 10.45 (s.d. = 4.84) semesters of formal classroom study. All participants had lived in Japan, with a range of 5-19 years (mean 3.5 years). Most continue to use Japanese professionally in the United States, as Japanese-language teachers, employees of Japanese organizations, or academics working on Japan-related topics. An additional four L2 subjects were tested but not included in the final analysis: three who were childhood bilinguals in English and a null argument language (Greek, Russian, Mandarin), and one with a low proficiency score.
5.2.4 Procedure

L1 participants were tested at a psycholinguistics laboratory at the University of Tokyo. Upon arrival in the lab, L1 participants were given a paper-and-pencil questionnaire in Japanese, eliciting the information on language background and English proficiency self-ratings reported above. Participants were then directed to a soundproof booth containing a Tobii 1750 eye-tracking computer for the main experiment. The eye-tracking computer was set up like a standard PC, minimizing the need for participants to familiarize themselves with its use. The experiment was programmed using E-Prime software, which ran on a laptop connected to the Tobii machine.

All participants completed a practice session consisting of sample items and a brief training session designed to familiarize them with the names and pictures of the four characters appearing in the test items (Takeshi, John, Kyoko, Susan). Once the practice session was complete, the experimenter answered any questions the participants had, then closed the door of the booth. During each item, participants saw a screen with four images, including two same-gender human figures, plus the direct object of the subordinate clause (DO1), and the setting of the story (see Section 5.2.2 for example item). This screen appeared two seconds before the audio began. Participants then heard a male Japanese native speaker reading the three sentences in the test item, followed by a one-second pause, and then the same speaker reading the Question. After the Question ended, participants used the mouse to click on their response. Three types of data were recorded for each participant: Interpretation (their answer to the Question), Reaction Time (from offset of Question to mouse click), and Eye-tracking for the duration of the trial (only analyzed for Test Sentence and Question). Most participants completed the
experiment in about an hour, including three built-in breaks. L1 participants were reimbursed ¥1000 for their time.

All but one L2 participant were tested at the CUNY Graduate Center; the remaining L2 participant was tested at the University of Tokyo lab where L1 data was collected. L2 participants followed an identical experimental procedure to that of L1 participants, with two differences: (i) a more detailed language-background questionnaire (in English), and (ii) a thirty-question proficiency test given after the main experiment. The questions for the proficiency test were culled from the listening comprehension sections of an old version of the Japanese Language Proficiency Test (JLPT), which the testing organization publishes and offers for sale to the general public (Japan Foundation and Japan Educational Exchanges and Services, 2012; 2012a; 2012b). The JLPT is given bi-annually and consists of five levels: N1 (advanced) through N5 (beginner). Test-takers choose their target level and take the test for that level only. For the proficiency measure in this study, I took ten listening-comprehension questions from each of the three highest levels (N1, N2, N3), for a total of 30 questions. For the main experiment, L2 participants sat in a sound-proof booth (similar to that found in the Tokyo lab) and completed the experiment on a Tobii 1760 eye-tracking computer, controlled by E-Prime software on a desktop outside the booth. L2 participants operated the Tobii computer like a normal PC. L2 participants took about an hour to complete the main experiment (similar to the L1 participants), while the proficiency test took them about thirty minutes. L2 participants were reimbursed $20 for their time.

5.2.5 Data Coding

Interpretation data was coded as either 1 for a topic-shift response (i.e., participants chose the non-subject referent) or 0 for a same-subject response (participants chose the subject of the
test sentence). Reaction time was measured straightforwardly in milliseconds from Question offset to mouse click. The eye-tracking data was collected at a 20 ms sampling rate; in other words, every 20 ms, the eye-tracker recorded which of the four pictures the participant was looking at. Each 20 ms sample was coded 1 or 0 for each of the following images depending on where the participant was looking at the time of the sample: TS Referent, SS Referent, Control-DO2, and Control-Setting.32 Each line of data therefore was coded with either a set of a single 1 and three 0s (reflecting the one image the participant was looking at) or a set of four zeros (indicating track loss). Trials with track loss were removed from analysis (33%).33

Eye-tracking data is presented below in two forms: time-course graphs for an informal overview, and inferential statistical models analyzing mean proportion of looks over entire constituents (DO2, V2, etc.). For the former, data was binned over 50 ms windows. Data was aggregated over participant and item for each condition (L1-waNull; L2-waNull; L1-waOvert; L2-waOvert, etc.) during each window with mean proportion of looks calculated for each image (TS Referent, SS Referent, Control). For example, calculating the mean proportion of looks for each image in the L1-waNull condition during the 0-50 ms window involved tallying all scores of “1” under each image type and dividing by the total number of trials (a total of 160, minus trials lost to track loss). The resulting means for the TS and SS referents were then plotted over time to create the time-course graphs.

Although I present mean proportion of fixations to the Topic-shift (TS) and Same-subject (SS) referents in time-course graphs to give a general overview of the data, all statistical analysis

---

32 Since the Control images revealed nothing of interest in preliminary analysis, they were grouped together under a single “Control” category in subsequent analysis.
33 One L2 participant was removed completely from the ET analysis due to high levels of track loss, leaving data from a total of twenty L1 participants.
of ET data was done using target advantage scores, with the TS referent as target (i.e., looks to the SS referent were subtracted from looks to the TS referent). This was done to provide further insight into the processing of sentences by participants. For example, if the target advantage score is close to zero, it suggests participants are vacillating between the two possible interpretations; simply providing the fixations to the TS referent or the SS referent would not capture this fact. Furthermore, instead of being binned over 50 ms windows as in the time-course graphs, the data for the statistical analysis was binned by syntactic unit (DO2, V2, etc.).

5.3 Predictions

This experiment tests two within-subjects conditions (Pronoun-Null/Overt and Particle-ga/wa), and one between-groups condition (Group-L1/L2). I discuss predictions for each condition below.

5.3.1 Pronoun Condition

Of the two within-subjects conditions, the Pronoun condition, reflecting the presence or absence of a pronominal subject in the matrix clause, is of primary concern. As discussed in Chapter 4, I predict that the addition of the supportive linguistic and visual context provided in these test items (as described above) will result in a consistent topic-shift interpretation for overt pronouns, as opposed to the ambiguous responses found in previous studies. Specifically, I predict:

• In the Null pronoun condition, L1 participants will show a consistent tendency (greater than chance) to choose a same-subject interpretation. In the Overt pronoun condition, L1 participants will show a consistent tendency (greater than chance) to choose a topic-shift interpretation (i.e., the non-subject referent).
5.3.2 Particle Condition

The Particle condition is a secondary condition, reflecting the choice of particle attached to the first noun in the subordinate clause: either the topic-marker –wa or subject-marker –ga. As mentioned above, this noun could potentially be marked with either, and it was unclear from a pilot study I conducted which would be preferable. Since it was likely that the choice of one over the other would have an effect on the results, both were included and Particle was added as a secondary condition. The literature on the exact functions on these two particles, and the distinction between them, is quite voluminous (for overview see Hinds, 1979; Heycock, 2008), but for the purposes of this paper we focus on a key syntactic contrast between the two that is particularly relevant for my Test Sentence: that a ga-marked NP can be the subject of a subordinate clause, whereas a wa-marked NP cannot (Kuno, 1973). Therefore, while ga-marked N can attach to the subordinate clause, wa-marked N cannot, and instead scopes over the entire sentence\(^\text{34}\):

\[\text{gaCondition} \quad \text{waCondition}\]

\[
\begin{array}{cccc}
\text{N-ga} & \text{DO1} & \text{V1} & \text{CONJ} \\
\text{Subordinate Clause} \\
\hline
\text{[PR or ø]} & \text{DO2} & \text{V2} \\
\text{Matrix Clause}
\end{array}
\]

\[
\begin{array}{cccc}
\text{N-wa} & \text{DO1} & \text{V1} & \text{CONJ} \\
\text{Subordinate Clause} \\
\hline
\text{[PR or ø]} & \text{DO2} & \text{V2} \\
\text{Matrix Clause}
\end{array}
\]

Previous studies have argued that when an overt topic and subject are in competition as possible antecedents of a null pronoun, Japanese native speakers choose the subject (Ueno & Kehler, \text{34} As mentioned elsewhere in this chapter, -wa has a contrastive function in addition to its function as a general topic marker. The contrastive reading is particularly common when both “subject” NPs in a bi-clausal sentence are marked with –wa. In order to minimize the chance of eliciting this contrastive reading, I marked the second NP (the Overt pronoun when present) with –ga instead of –wa.
However, in this experiment the two are not in competition, so the syntactic factor above is the main factor predicted to affect the results. Specifically, I predict:

- The *ga*-marked N will lead to more topic-shift interpretations than the *wa*-marked N, since the former but not the latter can be syntactically integrated into the subordinate clause, leaving the matrix clause free to have a different subject.

5.3.3 Group Condition

As discussed in Chapter 2, previous Interface Hypothesis (IH) experiments have found that advanced L2 speakers of Italian diverge from L1 speakers on precisely the kind of task used in this experiment: a forced choice interpretation of test sentences with a null/overt pronoun alternation (Sorace & Filiaci, 2006). Okuma (2012) found no differences between L1 and L2 speakers of Japanese on an IH-style task, but the L1 speakers also did not show a clear topic-shift preference for her specific items. While I argue in the previous chapter that the nature of the pragmatic aspect of overt pronoun interpretation differs in Japanese and Italian, there is no reason to suspect that the PCI involved in interpreting the overt pronoun in the former is less difficult to process than the GCI in the latter; speakers must still do a pragmatic computation in order to resolve the reference of a pronoun. If it truly is processing strains and/or structural complexity causing L1/L2 divergence in the previous IH studies, there is no reason to assume the L2 participants will be under any less processing strain in this experiment. I therefore predict that I will find similar L1/L2 divergence in my experiment to what was found in previous IH studies.

Previous research on L2 acquisition of Japanese particles suggests that the *ga* restriction on subjects of subordinate clauses is acquired relatively late among the various functions of *-wa* and *-ga*, certainly later than the topic function of *–wa*, which is acquired early (see Russell, 2005 for an overview), but as the L2 participants in this study are advanced, order of acquisition is not
particularly relevant. Russell (2005) found that L2 subjects who had lived in Japan and achieved a reasonable proficiency tended to produce more wa-marked sentences and fewer subordinate-clause –ga sentences ten years after returning home from Japan, but as his study was a production one (and subordinate clause structures are by nature complex and therefore relatively difficult to produce), it is unclear how the speakers would interpret these particles. Okuma (2012) featured a contrast between topic and subject pronoun antecedents in addition to the null/overt contrast, and found little difference between L1 and L2 participants in terms of the subject/topic distinction. Therefore, in the absence of strong evidence to the contrary, I predict that L2 speakers will not diverge from L1 participants on the Particle condition:

- L2 participants will choose topic-shift interpretations less often than L1 participants in the Overt condition, but the two groups will show similar interpretations in the Null condition. To the extent that L2 participants choose topic-shift interpretations, the effect will be mitigated by Particle in a way comparable to L1 participants.
CHAPTER 6

RESULTS

6.1 Introduction

In this section, I first review the interpretation data, including a detailed look at the individual results for L2 participants. I then investigate L1 and L2 processing patterns by examining reaction-time data, as well as eye-tracking data during both the Test Sentence and the Question. I end with an informal overview of the processing data separated into trials that follow the prediction for the Pronoun condition (With Prediction) and those that do not (Against Prediction). Overall, the interpretation data shows that L1 participants follow prediction for Pronoun and partially for Particle, while L2 participants do not follow prediction for either. The processing data for the most part conforms to the interpretation data, with more looks to the TS referent in Overt compared to Null conditions. Processing data also suggests for L1 participants, going against prediction for the Null causes the greatest amount of processing cost, suggesting that breaking the GCI in the Null is more difficult than calculating or canceling the PCI in the Overt. For L2 participants, going against prediction in the waOvert condition causes the most processing pressure, but this seems to be the result of violating the syntactic constraint imposed by the Particle. L2 participants do show some signs of processing the pronoun like L1 participants, but ultimately fall back on an English-like interpretation.
6.2 Interpretation Data

6.2.1 Overview: Interpretation Means

An informal examination of the means for the interpretation data reveals that L1 participants followed prediction for Pronoun and partially for Particle, whereas L2 participants did not follow prediction in either Pronoun or Particle\textsuperscript{35}:

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figures/fig2a.png}
\caption{Mean Proportion of Topic-shift Responses by Pronoun and Particle (L1, N=21)}
\end{figure}

\begin{figure}
\centering
\begin{subfigure}{0.49\textwidth}
\includegraphics[width=\textwidth]{figures/fig2b.png}
\caption{Mean Proportion of Topic-shift Responses by Pronoun (L1, N=21)}
\end{subfigure}
\begin{subfigure}{0.49\textwidth}
\includegraphics[width=\textwidth]{figures/fig2c.png}
\caption{Mean Proportion of Topic-shift Responses by Particle (L1, N=21)}
\end{subfigure}
\caption{Figures 2a, 2b, and 2c. Mean proportion of topic-shift responses by Pronoun and Particle (L1); by Pronoun only (L1) and by Particle only (L1)}
\end{figure}

\textsuperscript{35} Error bars indicate standard error. As explained in the next section, no outliers were removed from the interpretation data.
L1 participants chose a topic-shift referent for the Overt pronoun conditions in 76% of trials, but for the Null pronoun conditions in only 24% of trials, supporting an overall trend of topic-shift (TS) interpretation\(^{36}\) for Overts and same-subject (SS) interpretation\(^{37}\) for Nulls, following predictions for the Pronoun condition. L1 participants followed prediction for Particle only in the Null, with more TS responses in the \textit{gaNull} condition (36%) than \textit{waNull} condition (12%) but no differences between \textit{gaOvert} (76%) and \textit{waOvert} (76%). In sum, the L1 interpretation data seems to support prediction for Pronoun overall, and prediction for Particle only in the Nulls.

L2 participants, on the other hand, failed to follow prediction for either Pronoun or Particle, a result that I had predicted for the former condition but not the latter. Although there were more TS interpretations in the Overt conditions than the Null conditions for L2 participants, the overwhelming choice (> 50%) in both Pronoun conditions was the Same-subject referent. Furthermore, as reported in Section 6.3 below, the disparate rates in topic-shift between Overt and Null conditions may reflect bi-modality in the L2 data, with three prominent L2 outliers demonstrating L1-like means for TS rate in the Overt condition. The L2 data seems to show no effect of Particle, with almost identical TS rates in the two conditions (12% in \textit{wa} and 11% in \textit{ga}). Note that the disparity in the Particle condition across Group reflects not that L2 participants lack the syntactic restriction on topic-marked subjects in subordinate clauses (clearly they do, with low TS in \textit{waNull}), but that the L1 group allows topic-shift to happen in the \textit{gaNull} while the L2 group does not.

\(^{36}\) John in the example item given in Chapter 5.  
\(^{37}\) Takeshi in the example item given in Chapter 5.
Figures 3a, 3b, and 3c. Mean proportion of topic-shift responses by Pronoun and Particle (L2); by Pronoun only (L2) and by Particle only (L2)

6.2.2 Inferential Statistics: Interpretation Data

Inferential statistical analysis backs up the generalizations made from examining the descriptive statistics above. L1 and L2 means for each condition with confidence intervals appear below:
For the analysis, the interpretation data were coded either 1 (for topic-shift interpretation) or 0 (for same-subject interpretation), screened for outliers \((z > 3)^{38}\), and analyzed using the generalized linear mixed effects model function (glme4) in the R statistics software package (R Core Team, 2012; Bates, Maechler & Bolker, 2012). Random intercepts for participants and item were used to control for repeated measures effects.\(^{39}\) With random intercepts for participants and items in place, fixed effects were added one by one; each model was compared against previous models for goodness of fit using logLikelihood and the Akaike Information Criterion (AIC). I began by adding Group (L1/L2) as a fixed effect; the resulting model was significantly better than the null model with random effects only \(\chi^2(1) = 34.04, p < .001\). I next added Pronoun

---

\(^{38}\) Since removing outliers from categorical data in logistic regression entails removing all of a participant’s data for a specific condition, I chose a higher threshold for outliers here \((z > 3)\) than in the RT and ET analyses \((z > 2)\). There was only one outlier (in the L1-waOvert) condition, but removing their data made no difference in the model (including significant fixed effects and post-hoc contrasts). Therefore all data and analyses reported here include the outlier.

\(^{39}\) Random slopes models were considered but ultimately discarded as many models failed to converge.
(Overt/Null) and then Particle (ga/wa); both significantly improved the previous models ($\chi^2(1) = 525.2, p < .001; \chi^2(1) = 20.75, p < .001$). Adding the interaction of ParticleXGroup significantly improved the model over that with simple effects only ($\chi^2(1) = 12.37, p < .001$), and adding the interaction of PronounXGroup significantly improved the model yet again ($\chi^2(1) = 4.77, p = .029$). Hoping to capture the difference in Particle between Overt and Null conditions in the L1 data, I added a ParticleXPronoun interaction, which again significantly improved the model ($\chi^2(1) = 19.01, p < .001$). The three-way interaction (PronounXParticleXGroup) significantly improved the model with the three two-way interactions only ($\chi^2(1) = 11.9, p = .006$). The final model, with the three-way interaction, is presented below:

![Generalized linear mixed model fit by maximum likelihood (Laplace Approximation)]

---

Figure 5. R output for final model for interpretation data.
Post-hoc tests were conducted for the three-way interaction (Generalized Linear Hypotheses, adjusted Tukey contrasts). These reflect the lack of an effect of Particle in all but the L1 Null condition (\(waNull-gaNull; z = -7.505, p < .001\); there were no significant differences between \(waOvert\) and \(gaOvert\) in the L1 data (\(z = -.199, p = 1\)), or between \(wa\) and \(ga\) in the L2 Overt (\(z = .005, p = 1\)) or L2 Null (\(z = .873, p = .98\)) conditions. All other contrasts were significant, reflecting Pronoun and Group differences. Both L1 and L2 groups had significantly more TS responses in the Overt conditions compared to the Null (\(waOvert-waNull\) (L1): \(z = 15.48, p < .001\); \(gaOvert-gaNull\) (L1): \(z = 10.79, p < .001\); \(waOvert-waNull\) (L2): \(z = 6.01, p < .001\); \(gaOvert-gaNull\) (L2), \(z = 6.36, p < .001\)). However, the L2 TS rates for all conditions (both Overt and Null) were significantly lower than their L1 counterparts (\(waNull\): \(z = -3.01, p < .001\); \(gaNull\): \(z = -6.53, p < .001\); \(waOvert\): \(z = -7.02, p < .001\); \(gaOvert\): \(z = -7.1, p < .001\)). In sum, the inferential statistical analysis confirms the informal overview of the descriptives above, in that:

- Both groups show more TS responses in the Overt (L1-76%; L2-19%) than the Null (L1-24%; L2-4%) conditions.
- L1 participants give more TS responses than L2 participants in all conditions.
- L1 participants show an effect of Particle in the Null alone (\(waNull\)-12% TS; \(gaNull\)-36% TS). L2 participants show no effect of Particle.

6.2.3 L2 Individual Results, Proficiency and Language Background Survey Data

Looking more closely at the individual L2 means for Interpretation reveals some bimodality in the data: three L2 participants have TS rates close to the L1 mean.\(^40\) The chart on

\(^40\) All three participants were left in the final interpretation analysis, as their z-scores were below 3. The bimodality of the L2 data, however, is kept in mind throughout the interpretation of the
the left below shows the rate of topic-shift for Overt pronouns for individual L2 subjects. The lower line represents the L2 mean, and the upper line the L1 mean:

**Figures 6 & 7. Proportion of topic-shift for overt pronouns (individual results, L2) & Proportion of topic-shift for overt pronouns (individual results, L1).**

Overall, while the L2 interpretation data, unlike L1 data, does not follow predictions for Pronoun or Particle, three L2 participants (709, 817, 818) do display native-like behavior for the Overt condition. On the other hand, twelve participants (more than half) fall below the L2 mean, including six zeros. Only five (or 20% of) participants fall around the L2 mean. These results suggest that these three L2 participants have shifted to a native-like interpretation, whereas others have no topic-shift tendencies for overt pronouns at all, with a few in the middle. Notably, none of the participants are at chance, so it is unlikely that any are guessing. In comparison, the chart of individual L1 results on the right shows that most L1 participants are at or above the L1 mean, with a few outliers below.  

It is also important to note that while these three native-like L2-participants show native-like behavior on the Overt condition, they differ from L1 data. Taking these three participants out of the RT and ET data does not change the overall trends for this data in any noticeable way, so they are left in.

41 The lowest TS participant (906) is the outlier mentioned in the previous section.
participants in the Particle condition; specifically, they do not show any topic-shift in the Null, including gaNull.

The difference between these three L2 participants and the others is unclear. First of all, none of the three participants reported being childhood bilinguals in English and another null-argument language (such participants were dropped from the analysis), and only one reported learning another language in adulthood (Korean, from age 20, with low self-ratings for proficiency), so there should be no influence from an additional language. Furthermore, there is no correlation between proficiency score and rate of topic-shift in the Overt condition (r = .13, p = .58). L2 participants scored an average of 23.95/30 (80%) on the proficiency test (note that the test included items from intermediate through advanced levels, so that even the lowest scorers performed at least at intermediate level). Two out of three native-like L2-participants have high proficiency scores, but the third is below the mean (29, 29, and 21 out of 30). Also, many non-native-like participants have high proficiency scores (indeed, the only participant with a perfect proficiency score did no topic-shift).

### Figures 8 & 9. Correlation of proficiency and proportion of topic-shift response (L2) & Individual proficiency scores (L2).
I performed other correlations based on data from the L2 language-background survey (as reported in the Methodology chapter), and found no significant correlations between topic-shift rate and length of residence in Japan or current exposure (as measured by a self-rating scale). I found only two significant correlations with topic-shift rate in the Overt condition: age of onset of acquisition ($r = .51$, $p = .011$) and semesters of formal study ($r = .46$, $p = .02$):

**Figures 10 & 11. Correlation of age of onset of acquisition of Japanese and proportion of topic-shift responses (L2) & Correlation of semesters of formal study of Japanese and proportion of topic-shift responses (L2).**

The correlation between age and topic-shift in the Overt condition, while statistically significant, is somewhat problematic, since all participants (except one) began learning Japanese in middle-school or later, with a mean age of onset of 17.05 years. This puts the majority of participants beyond the age range typically considered for childhood bilingualism, with the lone exception not much younger (age 9). Furthermore, all of the participants who began studying Japanese at middle school age or younger (ages 9-14) did so in a foreign-language setting in English-speaking countries, with no exposure to Japanese in Japan until college. Among the three native-like L2 speakers, only two of the participants (709 and 817) began studying Japanese at a relatively young age (9 and 13 respectively); the third began studying at age 16, only a year below the mean. Also, four other participants started their classroom Japanese in middle school.
as well without showing native-like topic-shift in the Overt. The significance of the correlation of semesters of formal study and topic-shift is less problematic in theory, but somewhat surprising given that length of residence in Japan (and therefore amount of naturalistic vs. classroom L2 exposure) is not significant.

Ultimately, these correlations all suffer from the fact that while 14 out of 20 participants (70%) did do some topic-shift in the Overt condition, only 3 out of 20 participants (15%) did so at a rate comparable to L1 speakers. Therefore, the correlations above also reflect only the variation in TS rate among L2 participants, not the larger differences between L1 and L2 groups (i.e., it is less interesting if participants with a younger age of onset tend to have higher TS rates when those rates are as low as 6% or 12% or 25%, well under the L1 mean). It is the data for the three native-like participants that are of greater interest.

The data of the three native-like participants, summarized below, is somewhat heterogeneous, reflecting varied proficiency scores (817 was the 4th-lowest, 709 and 818 tied with a third participant for 2nd-highest), a slightly young age of onset, varied number of semesters (with two close to the mean, one high above it), varied length of time in Japan (with two quite below the mean and one quite above it), and varied current exposure ratings (one below the mean, two above). It is easy to see how this data might skew the overall correlations, since the two significant correlations (age and semesters) have extreme values for one of these subjects (709), whereas one potentially important factor (time in Japan) has low values for two out of the three (709 and 817). Furthermore, the one participant skewing all three of these values
is the one who was tested in Japan, a serious confound that cannot be controlled for statistically given the sample size of one.\textsuperscript{42}

**Table 10. Language Background Survey Results for Native-like L2 Participants**

<table>
<thead>
<tr>
<th>Subject</th>
<th>709</th>
<th>817</th>
<th>818</th>
<th>L2 Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>% Overt TS</td>
<td>75%</td>
<td>81.25%</td>
<td>75%</td>
<td>19%</td>
</tr>
<tr>
<td>Tested in</td>
<td>Japan</td>
<td>USA</td>
<td>USA</td>
<td>USA</td>
</tr>
<tr>
<td>Proficiency (out of 30)</td>
<td>29</td>
<td>21</td>
<td>29</td>
<td>23.95</td>
</tr>
<tr>
<td>Age of Onset</td>
<td>9</td>
<td>13</td>
<td>16</td>
<td>17.05</td>
</tr>
<tr>
<td>Semesters of Japanese</td>
<td>24</td>
<td>12</td>
<td>10</td>
<td>10.45</td>
</tr>
<tr>
<td>Time in Japan (mos.)</td>
<td>24</td>
<td>12</td>
<td>78</td>
<td>41.5</td>
</tr>
<tr>
<td>Current Exposure (out of 5)</td>
<td>3.56</td>
<td>2.25</td>
<td>3.38</td>
<td>2.67</td>
</tr>
</tbody>
</table>

This data suggests two possible routes for future study: (i) collecting more L2 data in Japan to allow a comparison between in-Japan and in-USA groups, (ii) collecting more L2 data from participants to allow a split between two age-of-onset groups (which will have to be somewhat arbitrary, since the traditional middle-school/puberty cut-off age seems not to be a factor here). For the rest of the data analysis (reaction time, eye-tracking), I keep these three participants in mind when interpreting results, but do not remove their data wholesale, leaving any undue influence of their data to be controlled through general screening for outlying trials.

### 6.3 Reaction Time (RT) Data

#### 6.3.1 Overview: RT Means

Examining means for the reaction time (RT) suggests little variation across conditions, with two possible effects worth exploring: a PronounXParticle interaction in the L1 data, and a

\textsuperscript{42} Taking this participant out completely does make Age and Semesters insignificant, while not quite bringing Time in Japan to significance.
main effect of Pronoun in the L2 data. If anything, the L1 graphs suggest a possible PronounXParticle interaction, in that the reaction times in the gaNull are about 400 ms slower than in waNull (and the longest overall), and that the waOvert condition is about 400 ms slower than gaOvert.

Figures 12a, 12b, & 12c. Reaction time by Pronoun and Particle (L1), by Pronoun (L1), and by Particle (L1).

---

43 Outliers replaced at $z > 2$. Error bars represent standard error.
The L2 charts suggest no effect of Particle (reflecting the interpretation data) but do suggest an effect of Pronoun. Perhaps most interestingly, a cursory look at the two groups suggests that L2 speakers may be much faster in the Null conditions than L1 speakers.

![Graph showing reaction times for L2 speakers](image1)

**Figures 13a, 13b, & 13c. Reaction time by Pronoun and Particle (L2), by Pronoun (L2), and by Particle (L2).**

### 6.3.2 Inferential Statistics: RT Data

Inferential statistical analysis does support the generalizations for the L2 and Group differences suggested above, but fail to find an interaction of PronounxParticle in the L1 data. To prepare
the data, I first searched for outliers on an individual-by-individual basis, calculating the value of two standard deviations above the mean for each participant and replacing any data points beyond that cutoff with that value.\textsuperscript{44} For the final data analysis, I log-transformed the data in order to achieve a normal distribution. The log-transformed RT data was then analyzed using the linear mixed effects model function in the R statistics software package. RT means (after replacement of outliers) with confidence intervals for both the raw and log-transformed data appear below:

\textbf{Figures 14a & 14b. Reaction time means (outliers replaced) & Reaction time means (outliers replaced, log transformed).}

As in the Interpretation data, random intercepts for participants and item were used to control for repeated measures effects. Fixed effects were then added one-by-one with each new model compared to the previous for goodness of fit by comparing logLikelihood and the Akaike Information Criterion (AIC). Adding Group (L1/L2) as a fixed effect did not significantly

\textsuperscript{44} Simply removing outliers with z-scores > 2 rather than replacing them did not change the analysis in any notable way. Removing outliers z > 3 does push the model with the PronounXParticle interaction into significance ($\chi^2(1) = 6.06, p = .048$), but I see no principled reason to set such a high cut-off for outliers.
improve the null model ($\chi^2(1) = .654, p = .42$), nor did a model with a fixed effect of Particle (ga/wa) ($\chi^2(1) = .07, p = .80$). I then tried a model with a fixed effect of Pronoun (Overt/Null), which proved significant ($\chi^2(1) = 19.31, p < .001$). I then added the interaction of PronounXGroup (reintroducing Group to the model as a simple effect as well), which significantly improved the model with Pronoun only ($\chi^2(2) = 15.68, p < .001$). Although the L1 charts above suggest an interaction of PronounXParticle, adding this interaction did not improve on the model with the PronounXGroup interaction ($\chi^2(2) = 2.43, p = .30$), nor did adding the three-way interaction improve the PronounXGroup model ($\chi^2(4) = 2.64, p = .62$). Therefore, the Particle effect in the L1 RT data does not seem to be significant. The final model is presented below:

```
Linear mixed model fit by maximum likelihood t-tests use
Satterthwaite approximations to degrees of freedom [merModLmerTest]
Formula: rt1 ~ pron + group + pron:group + (1 | id) + (1 | item)

AIC  BIC  logLik deviance df.resid
6285.8 6326.9  -3135.9 6271.8 2617

Random effects:
Groups   Name   Variance   Std.Dev.
    id   (Intercept)   0.26428   0.5141
   item   (Intercept)   0.01143   0.1069
 Residual            0.60308   0.7766
Number of obs: 2624, groups: id, 41; item, 8

Fixed effects:
                   Estimate   Std. Error    df t value Pr(>|t|)
(Intercept)       6.799e+00   1.221e-01 4.890e+01 55.676 < 2e-16 ***
pronovertnegativ  1.901e-02   4.237e-02 2.576e+03 0.449 0.653685
pronovertnegativ  1.496e-02   1.663e-01 4.380e+01 0.090 0.928692
pronovertnegativ  2.355e-01   6.066e-02 2.576e+03 3.882 0.000106 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

**Figure 15. R output for final model for reaction time data.**
Post-hoc tests (adjusted Tukey contrasts) revealed only one significant contrast: L2 speakers were significantly faster in the Null than the Overt condition ($z = 5.86$, $p < .001$). In sum, the inferential statistics reveal that:

- L2 participants were faster in the Null conditions (1298 ms) than the Overt conditions (1744 ms), an effect not found for the L1 participants (Overt = 1624 ms; Null = 1639 ms).

### 6.4 Eye-tracking (ET) Data: Test Sentence

#### 6.4.1 Overview: Test Sentence Time Course Data

In this section, I will first present and make informal observations of time-course graphs for the eye-tracking data.\(^{45}\) In the next section, I present a formal statistical analysis of the data during two time regions: the second direct object (DO2) and the second verb (V2), with a brief discussion of the data during the pronoun (PR) in the Overt conditions. Before examining the time-course graphs, recall the template for the Test Sentence, which is the second of three sentences in each mini-story. The graphs below appear with labels based on each constituent-based time region in the template (N = first Noun, DO1 = first direct object, V1 = first verb, CONJ = conjunction (aida, “while”), PR = pronoun, DO2 = second direct object, V2 = second verb).

**Test Sentence Template**

<table>
<thead>
<tr>
<th>N-wa/ga</th>
<th>DO1</th>
<th>V1</th>
<th>CONJ</th>
<th>[PR-ga or ø]</th>
<th>DO2</th>
<th>V2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Subordinate Clause</strong></td>
<td><strong>Matrix Clause</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Recall also that participants see four images on the screen: N, a second, same-gender human, the direct object of the first clause (DO1), and a picture of the setting (train, office, etc.). These images were labeled Same-subject Referent (SS), Topic-shift Referent (TS), Control (DO1), and Control (Setting) respectively. The latter two are not represented in the graphs, as they did not

\(^{45}\) Refer to Chapter 5 for information on ET data coding.
behave in a surprising way (increased looks to DO1 after participants hear it, and low looks to Setting overall) and have little bearing on the research questions.

In the graphs below, the lighter line shows the proportion of looks to the Same-subject referent (N in the Test Sentence template) while the darker line shows proportion of looks to the Topic-shift Referent (the other human). Vertical dotted lines represent mean temporal boundaries for each phrase by condition. We begin with the data for *gaNull*:

**Figure 16.** Eye-tracking data for Test Sentence (*gaNull*, L1).

**Figure 17.** Eye-tracking data for Test Sentence (*gaNull*, L2).
The data for the subordinate clause (unsurprisingly) did not vary much across condition (including Group), so I describe it here for gaNull as a representative sample. The ET data during the subordinate clause shows increased looks to the SS referent as participants hear it (during N) with a peak at DO1, followed by a decrease during V1 (as looks to the just-heard DO1 increase). In all conditions, there is an increase of looks to the TS referent during CONJ, as if participants are preparing themselves for the new clause signaled by the conjunction. In other words, both L1 and L2 participants are already showing increased TS looks before the matrix clause begins, regardless of Pronoun or Particle type, although this effect is much attenuated in the L2 Null.

The data for the gaNull condition during the matrix clause straightforwardly follows the interpretation data for both L1 and L2 participants. For L1 participants, TS looks continue to increase from CONJ into DO2 before peaking twice during V2, but only briefly overtake SS looks. This pattern is not out of line with the overall SS interpretation preference for L1 participants in this condition, an overall preference (64%) for SS, tempered by a 36% TS rate. In the L2 data, on the other hand, TS looks increase during CONJ as in the L1 data, but peak earlier (during DO2) and never come close to overtaking SS looks, which clearly dominate throughout the matrix clause. The L2 ET pattern reflects L2 reluctance to choose TS in the Null interpretation data, and L2 RT data, which was very quick in the Null, reflecting a lack of equivocation when choosing the SS referent. In sum, the gaNull ET data reinforce the interpretation and RT results in suggesting that L2 participants are much less willing than L1 participants to override the default Same-subject interpretation in the gaNull.

---

46 It is possible that there was an effect of Particle during the subordinate clause, but since this does not directly bear on the main research questions, I leave it for future analysis.
The data for *waNull* further emphasizes the differences between L1 and L2 participants in the Null condition. Even here, where L1 participants chose a TS referent only 12% of the time, there is a peak of TS looks near the onset of DO2 where looks to both referents come close to evening out, suggesting that the context given in these items does indeed support topic-shift to some extent. On the other hand, while L2 participants show a TS peak during DO2 as well, it is smaller than that in the L1, and looks to SS clearly dominate the L2 data throughout all of the matrix clause.

**Figure 18.** Eye-tracking data for Test Sentence (*waNull*, L1).

**Figure 19.** Eye-tracking data for Test Sentence (*waNull*, L2).
In the \textit{gaOvert} condition, the increase in TS looks during CONJ continues through the pronoun (PR), which seems to act as a kind of TS “booster.”

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure20.png}
\caption{Figure 20. Eye-tracking data for Test Sentence (gaOvert, L1).}
\end{figure}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure21.png}
\caption{Figure 21. Eye-tracking data for Test Sentence (gaOvert, L2).}
\end{figure}

In the L1 data, this increase in TS continues for the remainder of the matrix clause, overtaking SS looks in DO2 and V2. L2 participants too show increased TS looks during CONJ, PR, and DO2, but these never completely overtake SS looks, and decline sharply during V2. Overall, the
gaOvert data reflects the TS interpretation preference in the L1 data (76%), and the slight increase in TS interpretations compared to the Null in the L2 data (19%).

Finally, the waOvert, which was interpreted almost identically to the gaOvert by both L1 and L2 participants, shows similar patterns to the gaOvert data, although with some slight differences. Namely, there seem to be slightly less TS and more SS looks during the matrix clause in the L2 waOvert than the L2 gaOvert, even though both had a TS interpretation rate of 19%.

![Figure 22](image_url)  
*Figure 22. Eye-tracking data for Test Sentence (waOvert, L1).*

![Figure 23](image_url)  
*Figure 23. Eye-tracking data for Test Sentence (waOvert, L2).*
Overall, examination of the time-course graphs suggests processing differences that reflect the interpretation differences between groups. In particular, the ET data seems to underscore the certainty of L2 participants in the Null condition (i.e., they are not swayed by context) compared to L1 participants, who are more swayed by context in the gaNull (with accompanying processing cost, as seen in longer reaction times and more ambiguous ET data). The interaction of Particle and Pronoun in the L1 Null interpretation data may also be reflected in the increased TS looks in gaNull compared to waNull. Furthermore, while L2 group shows an increase in TS looks in the Overt conditions, suggesting they do have some kind of reaction to the pronoun, TS looks never overtake those to SS as they do in the L1 data. These tentative observations are investigated statistically below.

6.4.2 Inferential Statistics: ET Data

As described in Chapter 5, because of the importance of looks to both the TS referent and the SS referent, I used target advantage (TS looks-SS looks) as the dependent variable in the statistical analysis; TS was set as target in all conditions. Positive scores represent more looks to TS than SS; negative scores represent more looks to SS than TS, scores at or around zero represent vacillation between the two referents. I performed analyses of the target advantage data for each time region in the critical clause: Pronoun (PR), the direct object (DO2) and verb (V2) using the linear mixed effects model function in the R statistics software package, with subject and item as random intercepts. For each condition, outliers (z-score > 2) were removed before analysis.\footnote{Note that outliers were not removed in the time-course graphs.}

6.4.2.1 Pronoun (PR)
No significant differences were found during PR (the pronoun) in the Overt condition.

Target advantage means for the duration of the Pronoun appear below.\textsuperscript{48}

![Graph showing target advantage means for the duration of the Pronoun]

<table>
<thead>
<tr>
<th>Condition</th>
<th>L1 mean</th>
<th>L1 s.d.</th>
<th>L2 mean</th>
<th>L2 s.d.</th>
</tr>
</thead>
<tbody>
<tr>
<td>gaOvert</td>
<td>-.082</td>
<td>.182</td>
<td>-.082</td>
<td>.328</td>
</tr>
<tr>
<td>waOvert</td>
<td>-.092</td>
<td>.147</td>
<td>-.105</td>
<td>.284</td>
</tr>
</tbody>
</table>

**Figure 24 & Table 11.** Mean target advantage score during pronoun in graph & table form.

Examining looks during the utterance of the actual overt pronoun (*kare/kanojo*) removes the Pronoun condition from analysis (as there is no equivalent time period for Nulls), leaving Particle and Group conditions. Neither Group ($\chi^2(1) = .12, p = .73$) nor Particle ($\chi^2(1) = 1.77, p = .18$) significantly improved the null model when added, nor did the interaction ($\chi^2(1) = 2.37, p = .5$). Therefore, both L1 and L2 groups have the same ET patterns during the pronoun, with neither Group affected by Particle.

6.4.2.2 DO2

\textsuperscript{48} For the remainder of this chapter, the graphs represent means with 95% confidence intervals; the tables means with standard deviation.
The direct object of the matrix clause (DO2) provides the first chance to directly compare Null and Overt conditions. Mean target advantage during the utterance of DO2 is presented below.

![Graph](image)

**Figure 25 & Table 12. Mean target advantage score during DO2 in graph & table form.**

<table>
<thead>
<tr>
<th>Condition</th>
<th>L1 mean</th>
<th>L1 s.d.</th>
<th>L2 mean</th>
<th>L2 s.d.</th>
</tr>
</thead>
<tbody>
<tr>
<td>gaNull</td>
<td>.007</td>
<td>.273</td>
<td>-.240</td>
<td>.229</td>
</tr>
<tr>
<td>waNull</td>
<td>-.154</td>
<td>.303</td>
<td>-.218</td>
<td>.270</td>
</tr>
<tr>
<td>gaOvert</td>
<td>.206</td>
<td>.213</td>
<td>-.051</td>
<td>.215</td>
</tr>
<tr>
<td>waOvert</td>
<td>.031</td>
<td>.172</td>
<td>-.038</td>
<td>.317</td>
</tr>
</tbody>
</table>

Adding Group as a fixed effect significantly improved the model ($\chi^2(1) = 7.93, p = .005$), as did Pronoun ($\chi^2(1) = 63.6, p < .001$) and Particle ($\chi^2(1) = 10.89, p < .001$). Adding the interaction of ParticleXGroup further improved the model ($\chi^2(1) = 13.61, p < .001$). Neither the interaction of PronounXGroup ($\chi^2(1) = .12, p = .73$) or ParticleXPronoun ($\chi^2(1) = 1.16, p = .28$) significantly improved the model with the three simple effects only. The three-way interaction did not significantly improve the model with just the ParticleXGroup interaction ($\chi^2(3) = 6.28, p = .099$). The final model, with Group, Pronoun, Particle, and ParticleXGroup is presented below:
Linear mixed model fit by maximum likelihood t-tests use Satterthwaite approximations to degrees of freedom [merModLmerTest]
Formula: tadv ~ pron + group + part + part:group + (1 | id) + (1 | item)

<table>
<thead>
<tr>
<th>AIC</th>
<th>BIC</th>
<th>logLik</th>
<th>deviance</th>
<th>df.resid</th>
</tr>
</thead>
<tbody>
<tr>
<td>9557.4</td>
<td>9608.8</td>
<td>-4770.7</td>
<td>9541.4</td>
<td>4517</td>
</tr>
</tbody>
</table>

Random effects:

<table>
<thead>
<tr>
<th>Groups</th>
<th>Name</th>
<th>Variance</th>
<th>Std.Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>id</td>
<td>(Intercept)</td>
<td>0.019231</td>
<td>0.13868</td>
</tr>
<tr>
<td>item</td>
<td>(Intercept)</td>
<td>0.007723</td>
<td>0.08788</td>
</tr>
<tr>
<td>Residual</td>
<td></td>
<td>0.473358</td>
<td>0.68801</td>
</tr>
</tbody>
</table>

Number of obs: 4525, groups: id, 41; item, 8

Fixed effects:

| Estimate | Std. Error | df    | t value | Pr(>|t|) |
|----------|------------|-------|---------|---------|
| (Intercept) | 0.02702 | 0.05051 | 35.00000 | 0.535 | 0.596035 |
| pronover | 0.15238 | 0.02062 | 4506.00000 | 7.390 | 1.74e-13 *** |
| groupL2 | -0.24576 | 0.05338 | 57.00000 | -4.604 | 2.39e-05 *** |
| partwa | -0.15040 | 0.03035 | 4512.00000 | -4.955 | 7.49e-07 *** |
| groupL2:partwa | 0.15278 | 0.04139 | 4507.00000 | 3.692 | 0.000225 *** |

---

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

**Figure 26. R output for final model for target advantage data (DO2).**

The effects for Pronoun and Particle are straightforward, with more looks to the TS referent in Overt than Null, and in *ga* than *wa*, the latter in line with L1 data for *gaNull* vs. *waNull*. The effect for Group is also unsurprising, with L2 target advantage 25% lower than L1. Post-hoc tests of the ParticleXGroup interaction revealed that the difference between *wa* and *ga* conditions holds for L1 (z = -5.15, p < .001) but not L2 participants (z = .16, p = .99).

Furthermore, target advantage in the L1 *ga* condition was greater than in the L2 *ga* condition (z = -4.68, p < .001), but there was no difference between groups for *wa* conditions (z = -1.67, p = .32). In sum, the data during DO2 support the tentative observations made in the time-course data above, and support the interpretation data:
• More looks to TS referent in the Overt (target advantage: L1 = .119; L2 = -.135) compared to Null (L1 = -.074; L2 = -.229) in both groups.
• More looks to TS referent overall in the L1 group compared to the L2 group.
• An interaction of ParticleXGroup, in that there are more looks to TS in ga than wa for L1 participants only (L1 ga = .106; L1 wa = -.061).

6.4.2.3 V2

From the onset of the verb to the end of the sentence (V2), higher target advantage for Overt over Null continues for L1 speakers but neutralizes somewhat in the L2. Means with confidence intervals are presented below:

<table>
<thead>
<tr>
<th>Condition</th>
<th>L1 mean</th>
<th>L1 s.d.</th>
<th>L2 mean</th>
<th>L2 s.d.</th>
</tr>
</thead>
<tbody>
<tr>
<td>gaNull</td>
<td>-.087</td>
<td>.259</td>
<td>-.235</td>
<td>.237</td>
</tr>
<tr>
<td>waNull</td>
<td>-.165</td>
<td>.226</td>
<td>-.221</td>
<td>.170</td>
</tr>
<tr>
<td>gaOvert</td>
<td>.086</td>
<td>.274</td>
<td>-.154</td>
<td>.238</td>
</tr>
<tr>
<td>waOvert</td>
<td>.123</td>
<td>.224</td>
<td>-.170</td>
<td>.314</td>
</tr>
</tbody>
</table>

Figure 27 & Table 13. Mean target advantage score during V2 in graph & table form.
Adding Group as a fixed effect significantly improved the null model ($\chi^2(1) = 9.91, p = .002$).
Adding Pronoun improved the model with Group only ($\chi^2(1) = 68.23, p < .001$), but further
adding Particle did not ($\chi^2(1) = .06, p = .81$). Adding an interaction of PronounXGroup
significantly improved the model with Group and Pronoun only ($\chi^2(1) = 22.04, p < .001$). Adding
further interactions (including the three-way interaction) did not significantly improve the model.

The final model is presented below:

```
Linear mixed model fit by maximum likelihood t-tests use Satterthwaite approximations to degrees of freedom [merModLmerTest]
Formula: tadv ~ pron + group + pron:group + (1 | id) + (1 | item)
AIC  BIC  logLik deviance df.resid
15342.9 15391.3  -7664.4 15328.9    7423

Random effects:
  Groups   Name   Variance  Std.Dev.
         id (Intercept) 0.025933  0.1610
         item (Intercept) 0.003261  0.0571
   Residual            0.453976  0.6738
Number of obs: 7430, groups: id, 40; item, 8

Fixed effects:
                     Estimate  Std. Error   t value     Pr(>|t|)
(Intercept)        -0.122200   0.044422   -2.7511     0.00841 **
pronovert          0.210970   0.023209     9.0900      < 2e-16 ***
groupL2            -0.106990   0.055574    -1.9253     0.06041 .
pronovert:groupL2  -0.148500   0.031605    -4.6989     2.66e-06 ***

---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

**Figure 28. R output for final model for target advantage data (V2).**

Looks to the target remain higher in the Overt compared to the Null in all conditions. The effect
of Particle in the L1 data, on the other hand, has disappeared. Post-hoc tests reveal that target
advantage remains higher in the Overt compared to the Null for both L1 ($z = 9.1, p < .001$) and
L2 ($z = 2.9, p = .015$) participants, but that target advantage is higher for L1 than L2 participants
in the Overt conditions ($z = -4.55, p < .001$). Furthermore, the L2 Overt condition is not
significantly different from the L1 Null condition \( (z = -0.8, p = 0.84) \), meaning that the L2 Overt data patterns with the L1 Null data. In sum, during V2:

- Looks to the TS referent remain higher in the Overt (target advantage L1 = 0.104; L2 = -0.187) than Null (L1 = -0.126; L2 = -0.228) conditions for both groups.

- Within the Overt condition, L1 participants fixate on the TS referent to a greater extent than L2 participants. The L2 *Overt* target advantage score (-0.187) is not significantly different from the L1 *Null* score (-0.126).

- The effect of Particle in the L1 group disappears. There remains no effect of Particle in the L2 group.

6.4.3 Summary

To sum up, the ET data during the matrix clause reveals group differences that reflect the interpretation and reaction time data. Although both groups start out with similar ET patterns during the pronoun (PR), they diverge during the direct object (DO2) and continue to differ throughout the verb to the end of the sentence (V2). Although L2 participants do fixate on TS more often in the Overt condition than the Null, they do so significantly less than L1 participants in both DO2 and V2 (in both the Overt and the Null), with consistently negative target advantage in all conditions. This suggests that L2 participants are not completely immune to the pronoun, just that they react differently to it than L1 participants. L2 participants are also less likely than L1 participants to let context override the default same-subject bias for the Null. Furthermore, L1 participants show an effect of Particle during DO2, with more TS looks in *ga* than *wa*; L2 participants show no such effect. This Particle effect disappears by the end of the Test Sentence (during V2).
6.5 Eye-tracking Data: Question

6.5.1 Overview: Question Time Course Data

After hearing each test item—Introduction, Test, and Concluding Sentences—the participants hear a Question asking them who performed the activity in the matrix clause. I divided the analysis of the ET data for the Question into two time frames: during the Question itself (Q) and After the Question (AQ), the period of time from the offset of the Question until participants click their answer. The length of the AQ period is naturally variable, since it is determined by reaction time for each trial. In the graphs below, mean reaction time for each condition is included for reference, but the further to the right end of the graph, the fewer trials are represented. The Question itself is divided into its constituent parts on the graphs, but these parts are combined in the statistical analysis:

<table>
<thead>
<tr>
<th>Question Template</th>
<th>After Question Period</th>
</tr>
</thead>
<tbody>
<tr>
<td>WHO DO2 V2 ka</td>
<td>AQ</td>
</tr>
<tr>
<td>Q</td>
<td></td>
</tr>
</tbody>
</table>

WHO refers to the question word and attached subject particle (dare-ga in Japanese), DO2 and V2 are the same as in the Test Sentence, and are followed by ka, the Japanese question particle. As in Section 6.4, I begin with a descriptive account of the time course graphs before reporting a more formal statistical analysis; the lighter line shows the proportion of looks to the Same-subject referent and the darker line shows proportion of looks to the Topic-shift referent. In the gaNull condition, both groups fixate more on the SS referent (the preferred response for both groups) from the start of the Question and throughout the After Question period. However, looks to SS are much higher in the L2 group, and the L1 group shows increased looks to the TS referent in the AQ period, in line with their increased responses choosing the TS referent (36% in L1 vs. 3% in L2).
For the Question and After-Question periods, separating the trials by ultimate interpretation response is quite useful. An overview of the data thus divided is given here, with greater detail reported in Section 6.6. Data is divided into trials in which participants answer With Prediction (WP) for Pronoun and those in which they answer Against Prediction (AP) for Pronoun.

Indeed, separating the L1 gaNull data into With Prediction (SS interpretation) and Against Prediction (TS interpretation) trials reveals that the increased looks to TS at the expense
of SS in the combined graph above reflect the Against Prediction (AP) trials, while the With Prediction (WP) trials more resemble the L2 combined data (and the L1 \textit{waNull} data).

![Figure 31. Eye-tracking data for Question (gaNull-Against Prediction, L1).](image1)

![Figure 32. Eye-tracking data for Question (gaNull-With Prediction, L1).](image2)

The L1 \textit{gaNull} Against Prediction trials represent the longest reaction time of any condition (3250 ms, compared to 1467 ms With Prediction), which is reflected in the time-course data. After a long period of increased looks to TS during and after the question, L1 participants who have not already answered the question appear to enter a period of vacillation for about 750 ms approaching the mean RT, suggesting that the increased processing cost indicated by the
increased reaction time reflects reluctance to choose against prediction for the Null, a reluctance that is perhaps both conscious and overcome consciously. L2 speakers, on the other hand, almost never choose against prediction in the Null conditions in the first place (in fact, there is so little Against Prediction data for L2 participants that it is difficult to analyze and thus not included here). It seems as if answering against prediction in the Null is so costly in terms of processing that L2 participants don’t do it, a fact we will explore further in the Discussion chapter.

Figure 33. Eye-tracking data for Question (waNull, L1).

Figure 34. Eye-tracking data for Question (waNull, L2).
In keeping with its more straightforward interpretation data (12% TS in L1, 4% in L2), the waNull ET data shows a clear looking preference to the SS referent for both L1 and L2 (see Figures 33 & 34 above). This trend becomes even clearer when examining With Prediction trials in both groups. The Against Prediction trials are not included here, since both groups have little data (in fact, I argue in section 6.6 that the AP waNull trials largely represent error for both groups):

![Figure 35. Eye-tracking data for Question (waNull-With Prediction, L1).](image)

![Figure 36. Eye-tracking data for Question (waNull-With Prediction, L2).](image)
In the two Overt conditions, both Groups fixate preferentially on their preferred interpretation choice from the onset of the Question: L1 on the TS referent (although with some vacillation around mean RT) and L2 participants on the SS referent. In keeping with the interpretation data, there do not seem to be any differences in Particle in the overall pattern of the L1 data, as opposed to the Null:

**Figure 37.** Eye-tracking data for Question (waOvert, L1).

**Figure 38.** Eye-tracking data for Question (waOvert, L2).
Examining only the trials in which L1 participants follow prediction and choose a TS referent (~76% both gaOvert and waOvert) somewhat tempers but does not completely remove the vacillation around mean RT (which remains greater in waOvert then gaOvert) This vacillation is not found in the L1 With Prediction trials for the Null conditions; removing the Against Prediction trials removes all ambivalence in look preference. This disparity suggests that
even though L1 participants clearly fixate on their eventual answer from the start of the question, they may do a quick “double-check” when it comes time to click in the Overt conditions:

![Figure 41. Eye-tracking data for Question (gaOvert-With Prediction, L1).](image)

![Figure 42. Eye-tracking data for Question (waOvert-With Prediction, L1).](image)

L2 participants chose a TS interpretation in 19% of trials in both Overt conditions; however, their ET data in the WP trials does not quite match the L1 WP data. The overall pattern is similar in the gaOvert condition (especially considering the fewer trials make for grainier data), but the waOvert is clearly different, with no obvious looking preference throughout. This
disparity suggests that despite the interpretation data, and despite the ET data during the Test Sentence, L2 participants may show some sensitivity to Particle after all. Note the waOvert but not the gaOvert prediction for Pronoun (TS) and Particle (SS) are at odds with each other, potentially creating extra processing pressure. Specifically, in the WP trials, L2 participants are choosing the context-supported TS interpretation despite the syntactic constraint against TS by the Particle.

Figure 43. Eye-tracking data for Question (gaOvert-Without Prediction, L2).

Figure 44. Eye-tracking data for Question (waOvert-Without Prediction, L2).
When choosing Against Prediction, L1 participants show some vacillation during the Question but eventually fixate on their ultimate choice (SS) in the AQ period, although TS looks remain relatively high in the waOvert. This vacillation is not as strong as in the AP gaNull. L2 participants, on the other hand, show less vacillation, as the AP is their preferred choice in the Overt.

Figure 45. Eye-tracking data for Question (gaOvert-Against Prediction, L1).

Figure 46. Eye-tracking data for Question (waOvert-Against Prediction, L1).
Figure 47. Eye-tracking data for Question (gaOvert-Against Prediction, L2).

Figure 48. Eye-tracking data for Question (waOvert-Against Prediction, L2).

Note that the two groups showed opposite patterns in the *waOvert*: while L2 participants showed vacillation when answering With Prediction (choosing the context-driven interpretation of the Pronoun over the syntactic constraint of the Particle), L1 participants show vacillation when answering *Against* Prediction (choosing the syntactic constraint of Particle over the context-driven interpretation of the Pronoun).
In sum, visual inspection of the ET data during and after the Question:

- conforms with the interpretation and RT data and the Group differences found therein.
- suggests severe processing cost during the L1 gaNull condition when choosing Against Prediction.
- reveals less processing cost in Overt WP/AP compared to Null AP for L1 participants
- shows differences between L1 and L2 groups in processing the Question in the waOvert condition, with more difficulty for choosing WP (TS) in L2 and AP (SS) in L1.

6.5.2 Inferential Statistics: ET Data-Question

As with the ET data during the Test Sentence, I used Target Advantage as the dependent variable in the statistical analysis for the ET data for the Question. I performed analyses of the data for the Question itself and for the time between Question offset and mouse click (AQ) using the linear mixed effects model function in the R statistics software package, with subject and item as random intercepts. For each condition, outliers (z-score > 2) were removed before analysis. Only the combined data (not separated by With/Against Prediction) is analyzed in this section due to the uneven data that arises when the data is separated, although descriptives by WP/AP trials will be examined in further detail in section 6.6.

6.5.2.1 Question

Target advantage means during the Question (in the graph below) reflect the overall difference between L1 and L2 participants: more looks to the TS referent in the Overt than Null for the former, no differences for the latter.
Adding Group as a fixed effect significantly improved the null model ($\chi^2(1) = 18.31$, $p < .001$); further adding Pronoun ($\chi^2(1) = 376.47$, $p < .001$) and then Particle ($\chi^2(1) = 6.38$, $p = .012$) caused significant improvement over previous models. A model including the interaction of PronounXGroup was significantly better than the model with simple effects only ($\chi^2(1) = 88.94$, $p < .001$); further adding the ParticleXGroup interaction improved the model with the PronounXGroup interaction ($\chi^2(1) = 4.53$, $p = .03$). No other interactions (including the three-way interaction) further improved the model.

**Table 14.** Mean target advantage during Question in graph & table form. 

<table>
<thead>
<tr>
<th>Condition</th>
<th>L1 mean</th>
<th>L1 s.d.</th>
<th>L2 mean</th>
<th>L2 s.d.</th>
</tr>
</thead>
<tbody>
<tr>
<td>gaNull</td>
<td>-.181</td>
<td>.220</td>
<td>-.375</td>
<td>.194</td>
</tr>
<tr>
<td>waNull</td>
<td>-.194</td>
<td>.279</td>
<td>-.337</td>
<td>.283</td>
</tr>
<tr>
<td>gaOvert</td>
<td>.078</td>
<td>.254</td>
<td>-.225</td>
<td>.202</td>
</tr>
<tr>
<td>waOvert</td>
<td>.166</td>
<td>.309</td>
<td>-.232</td>
<td>.259</td>
</tr>
</tbody>
</table>
Linear mixed model fit by maximum likelihood t-tests use Satterthwaite approximations to degrees of freedom [merModLmerTest]
Formula: tadv ~ group + pron + part + part:group + pron:group + (1 | id) + (1 | item)

<table>
<thead>
<tr>
<th>AIC</th>
<th>BIC</th>
<th>logLik</th>
<th>deviance</th>
<th>df.resid</th>
</tr>
</thead>
<tbody>
<tr>
<td>36520.8</td>
<td>36590.2</td>
<td>-18251.4</td>
<td>36502.8</td>
<td>16537</td>
</tr>
</tbody>
</table>

Random effects:
<table>
<thead>
<tr>
<th>Groups Name</th>
<th>Variance</th>
<th>Std.Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>id (Intercept)</td>
<td>0.026438</td>
<td>0.16260</td>
</tr>
<tr>
<td>item (Intercept)</td>
<td>0.003947</td>
<td>0.06282</td>
</tr>
<tr>
<td>Residual</td>
<td>0.527151</td>
<td>0.72605</td>
</tr>
</tbody>
</table>

Number of obs: 16546, groups: id, 40; item, 8

Fixed effects:
| Estimate | Std. Error | df   | t value | Pr(>|t|) |
|----------|------------|------|---------|---------|
| (Intercept) | -2.320e-01 | 4.523e-02 | 4.800e+01 | -5.129 | 5.18e-06 *** |
| groupL2    | -1.234e-01 | 5.539e-02 | 4.800e+01 | -2.227 | 0.030655 *   |
| pronover   | 3.350e-01  | 1.644e-02 | 1.651e+04 | 20.381 | < 2e-16 ***  |
| partwa     | 5.518e-02  | 1.668e-02 | 1.654e+04 | 3.309  | 0.000938 *** |
| groupL2:partwa | -4.865e-02 | 2.287e-02 | 1.654e+04 | -2.127 | 0.033400 *   |
| groupL2:pronover | -2.149e-01 | 2.269e-02 | 1.651e+04 | -9.471 | < 2e-16 ***  |

---
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Figure 50. R output for final model for target advantage data (Question).

Post-hoc tests revealed that the Overt trials had significantly higher target advantage scores than Null trials for both L1 (z = 20.32, p < .001) and L2 (z = 7.67, p < .001) groups; however, target advantage was significantly higher for L1 participants than L2 participants in both Overt (z = -6.72, p < .001) and Null (z = -2.77, p = .02) conditions, with no significant difference between L2 Overt and L1 Null conditions (z = -.56, p = .93). Target advantage was higher in wa conditions than ga conditions for L1 (z = 2.94, p = .013) but not L2 (z = .34, p = .98) participants. In sum, during the Question:

- Both groups have more TS looks in Overt (target advantage: L1 = .122; L2 = -.281) than Null (L1 = -.187; L2 = -.356).
• L1 participants have more TS looks than L2 in both Overt and Null (note L2 Overt target advantage is still negative, even though it is larger than the Null).

• Target advantage is higher in wa than ga for L1 participants (wa = -.014; ga = -.052), despite the fact during the Test Sentence the opposite was true. As suggested by the means graph at the beginning of this section, and as we will see in the AQ data, this effect is most likely due to the Overt condition, not the Null.

6.5.2.2 After Question (AQ)

![Graph and Table]

<table>
<thead>
<tr>
<th>Condition</th>
<th>L1 mean</th>
<th>L1 s.d.</th>
<th>L2 mean</th>
<th>L2 s.d.</th>
</tr>
</thead>
<tbody>
<tr>
<td>gaNull</td>
<td>-.281</td>
<td>.309</td>
<td>-.678</td>
<td>.215</td>
</tr>
<tr>
<td>waNull</td>
<td>-.456</td>
<td>.247</td>
<td>-.559</td>
<td>.253</td>
</tr>
<tr>
<td>gaOvert</td>
<td>.272</td>
<td>.490</td>
<td>-.405</td>
<td>.320</td>
</tr>
<tr>
<td>waOvert</td>
<td>.332</td>
<td>.379</td>
<td>-.526</td>
<td>.329</td>
</tr>
</tbody>
</table>

**Figure 51 & Table 15.** Mean target advantage during AQ in graph & table form.

Target advantage means in the period from Question offset to mouse click (AQ) continue to reflect the overall differences between L1 and L2 participants, with greater target advantage for the Overt than Null in L1 but not L2. However, the three-way interaction of
PronounXParticleXGroup becomes significant, reflecting a newfound difference in Particle for L2 participants. Adding Group as a fixed effect significantly improved the null model ($\chi^2(1) = 30.94, p < .001$); adding Pronoun further improved the model with Group only ($\chi^2(1) = 800.6, p < .001$). However, adding Particle did not significantly improve the model ($\chi^2(1) = .68, p = .57$). An interaction of PronounXGroup significantly improved the model with Group and Pronoun alone ($\chi^2(1) = 307.4, p < .001$). Looking at the means, I decided to add an interaction of ParticleXGroup as well, which significantly improved the model with just the PronounXGroup interaction ($\chi^2(1) = 8.42, p = .015$). Adding the three-way interaction further improved the model ($\chi^2(1) = 136.68, p < .001$). The final model appears below:

| Linear mixed model fit by maximum likelihood t-tests use Satterthwaite approximations to degrees of freedom [merModLmerTest] Formula: tadv ~ pron * part * group + (1 | id) + (1 | item) |
|-----------------------------------------------|
| AIC  | BIC   | logLik | deviance | df.resid |
| 28082.8 | 28164.9 | -14030.4 | 28060.8 | 12763 |

Random effects:

<table>
<thead>
<tr>
<th>Groups</th>
<th>Name</th>
<th>Variance</th>
<th>Std.Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>id</td>
<td>(Intercept)</td>
<td>0.054441</td>
<td>0.23333</td>
</tr>
<tr>
<td>item</td>
<td>(Intercept)</td>
<td>0.003968</td>
<td>0.06299</td>
</tr>
<tr>
<td></td>
<td>Residual</td>
<td>0.520226</td>
<td>0.72127</td>
</tr>
</tbody>
</table>

Number of obs: 12774, groups: id, 41; item, 8

Fixed effects:

|                | Estimate  | Std. Error | df      | t value | Pr(>|t|) |
|----------------|-----------|------------|---------|---------|---------|
| (Intercept)    | -2.632e-01| 5.957e-02  | 5.300e+01| -4.418  | 4.92e-05 *** |
| pronover        | 4.600e-01 | 2.638e-02  | 1.275e+04| 17.436  | < 2e-16 *** |
| partwa         | -2.311e-01| 2.702e-02  | 1.275e+04| -8.554  | < 2e-16 *** |
| groupL2        | -4.366e-01| 7.836e-02  | 4.700e+01| -5.572  | 1.19e-06 *** |
| pronover:partwa| 3.519e-01 | 3.795e-02  | 1.275e+04| 9.272   | < 2e-16 *** |
| pronover:groupL2| -1.626e-01| 3.655e-02  | 1.274e+04| -4.450  | 8.68e-06 *** |
| partwa:groupL2 | 3.859e-01 | 3.765e-02  | 1.274e+04| 10.251  | < 2e-16 *** |
| pronover:partwa:groupL2 | -6.064e-01| 5.196e-02  | 1.275e+04| -11.671 | < 2e-16 *** |

---

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Figure 52. R output for final model for target advantage data (After Question).
Post-hoc testing revealed higher target advantage in the Overt compared to Null for both groups in the ga conditions (gaOvert-gaNull (L1): $z = 17.44$, $p < .001$; gaOvert-gaNull (L2): $z = 11.75$, $p < .001$), but for L1 only in the wa conditions (waOvert-waNull (L1): $z = 29.51$, $p < .001$; waOvert-gaOvert, $z = -4.14$, $p < .001$). L2 participants had lower target advantage scores than L1 participants in all conditions (gaNull, $z = -5.57$, $p < .001$; gaOvert, $z = -7.72$, $p < .001$; waOvert, $z = -10.50$, $p < .001$), except for waNull, where there were no differences ($z = .65$, $p = .997$). In sum:

- Both groups have higher target advantage scores in the Overt than the Null in the ga conditions (L1 gaOvert = .272; L1 gaNull = -.281; L2 gaOvert = -.405; L2 gaNull = -.678), but only L1 participants do so in the wa conditions (L1 waOvert = .332; L1 waNull = -.456; L2 waOvert = -.526; L2 waNull = -.559), reflecting vacillation for L2 participants in the waOvert (reflecting WP trials).
- L1 participants fixate more on the TS referent than L2 participants in all conditions. (Again, the target advantage score in L2 gaOvert is still very low, even though it is significantly higher than the Null).
- The L1 data for Particle conforms with interpretation data for the Null, with more TS looks in ga (L1 gaNull = -.281) than wa (L1 waNull = -.456), but not for the Overt, where there is a higher target advantage for wa (L1 waOvert = .332) than ga (L1 gaOvert =
.272), despite the fact that \textit{waOvert} and \textit{gaOvert} have almost identical interpretation patterns (both at 76\% TS).

6.5.3 \textit{Summary: Question and After Question Periods}

Overall, inferential statistical analysis of the data in the Q and AQ periods supports the observations made by examining the time-course graphs. The ET data reflects the interpretation data, with more target advantage scores in the Overt than the Null for both groups, but with higher target advantage scores for L1 than L2 overall. For the L1 participants, there are more looks to TS in \textit{gaNull} than \textit{waNull}, also in line with the interpretation data. For both the L1 and L2 groups, there is some divergence from the interpretation data in the Overt condition. Although there was no effect of Particle in interpretation, there was an effect in the ET data: more vacillation in \textit{waOvert} than \textit{gaOvert} in L2 and higher TS rates in \textit{waOvert} than \textit{gaOvert} in L1. While these effects were unexpected based on the interpretation data, they are somewhat unsurprising, as \textit{waOvert} is a syntactically clunky construction, with a topic scoping over a matrix clause with another overt subject. Furthermore, dividing the Question into WP and AP trials reveals that L2 participants show more vacillation in \textit{waOvert} when overcoming the syntactic Particle constraint against topic-shift in WP trials, while L1 participants show more vacillation when choosing SS, choosing with the Particle constraint against the context-imposed PCI. In the next section, we further explore the data as divided into trials by ultimate interpretation response.

6.6 \textit{Choosing With and Against Prediction}

When comparing the WP/AP processing data, therefore, it is first necessary to recall how the two groups differ in interpreting the pronoun in the first place. L1 participants chose against the predictions for the Pronoun condition 36\% of the time in the \textit{gaNull} (TS instead of SS), 24\%
of the time in the two Overt conditions (SS instead of TS), and only 12% of the time in \textit{waNull} (TS instead of SS). L2 participants chose against prediction 81% of the time in the Overt conditions, and only 4% of the time in the Null conditions. L1 participants, therefore, overall follow prediction, so that choosing against prediction is choosing against their overall preferred interpretation for that condition. L2 participants prefer a Same-subject referent in the Overt conditions, meaning they prefer a response that is against prediction for Pronoun in the first place. In the Null conditions, on the other hand, they again prefer Same-subject, meaning they tend to follow prediction; there is so little AP data in the Null conditions (with the except of L1 \textit{gaNull}) that it is difficult to analyze the AP data alone. In reviewing the data for WP and AP trials, therefore, I consider descriptives only, due to the unbalanced nature of the data, which makes me reluctant to use inferential statistics (in the most extreme example, almost all of the L2 Null interpretation data is in WP trials, with only a handful in AP, representing data from only 25% of participants).\footnote{In the graphs below, AP is labeled “0” and WP “1”; bars represent 95% confidence intervals.}

6.6.1 Reaction Time by WP/AP Trials

Examining the reaction time data for L1 participants, we see (i) reaction times are faster in the WP trials than AP trials for all conditions, (ii) large variability in AP trials, especially in the Null conditions, and (iii) \textit{gaNull} with a higher mean than the other three conditions (note that \textit{waNull} has the lowest amount of AP data, followed by the two Overt conditions, with \textit{gaNull} having the most).
In sum, the L1 reaction time data suggests greater processing effort when choosing against prediction in all conditions, as seen in overall slower times in AP trials, as well as greater variability. However, the *gaNull* condition in particular seems to entail great processing difficulty, with high RT and variability, despite having the most data of all L1 AP conditions.

Comparing L1 and L2 means (see Figures 54 & 55) further reinforces the group differences in interpretation: (i) L2 data is slowest not in AP trials but in Overt WP trials, especially *waOvert*, (ii) L2 participants show greatest variability when choosing topic-shift (Overt WP, Null AP), and (iii) L2 participants are fast in the Null regardless of their ultimate answer.
Figures 54 & 55. Mean reaction time for trials answered with prediction for Pronoun & against prediction for Pronoun.

These results clearly reflect the patterns found in the interpretation and ET data, suggesting that L2 participants expend their greatest processing effort when choosing topic-shift, especially in the waOvert (whereas L1 participants expend their greatest processing effort choosing against prediction in gaNull).²⁵⁻⁰

6.6.2 Eye-tracking Data by WP/AP Trials: L1 Group

Examining the time-course ET data during the Test Sentence shows three patterns for L1 participants: for waNull and gaOvert, the ET data looks almost identical regardless of final choice (especially considering that the data is grainier in AP trials). Note that these are the two conditions in which prediction for Particle and Pronoun match up (SS in waNull and TS in gaOvert):

²⁵ The low amount of AP interpretations (~4%) in the Null, taken with these fast reaction times, suggests that the L2 AP Null trials may be the result of pure error (lack of attention, clicking the wrong image by accident) rather than a principled decision (as I will also suggest is the case for L1 AP waNull).
Figure 56. Eye-tracking data for Test Sentence (waNull-Against Prediction, L1).

Figure 57. Eye-tracking data for Test Sentence (waNull-With Prediction, L1).
Figure 58. Eye-tracking data for Test Sentence (gaOvert-Against Prediction, L1).

Figure 59. Eye-tracking data for Test Sentence (gaOvert-With Prediction, L1).

For waOvert, the ET data looks slightly different in WP and AP trials, although with the same overall trend by V2. Recall that in the waOvert condition, the prediction for Pronoun (TS) is at odds with the prediction for Particle (SS), based on syntactic condition that topic-marked NPs
cannot be subjects of subordinate clauses, and hence must scope over the entire sentence (recall also the slight vacillation in the AP waOvert during the Q and AQ periods).\textsuperscript{51}

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure60.png}
\caption{Eye-tracking data for Test Sentence (waOvert-\textit{With} Prediction, L1).}
\end{figure}

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure61.png}
\caption{Eye-tracking data for Test Sentence (waOvert-\textit{Against} Prediction, L1).}
\end{figure}

Finally, gaNull also shows some differences in AP vs. WP:

\textsuperscript{51} For waOvert there also seems to be some odd fixation patterns in the subordinate clause (N, wa, DO1, V2, CONJ) in AP compared to WP. I’m not sure what this would be, since it happens well before the pronoun, and even if there were some biasing content in certain items, it should show up in waNull as well, as the two conditions are identical up to that point.
There are slightly high TS looks in WP trials (especially compared to \textit{waNull}), but overall SS dominates. In AP trials, TS looks clearly overtake SS in a similar way to WP Overt conditions. In \textit{gaNull}, the prediction for Pronoun is SS, but as the subject-marking on the first N allows the first N to be syntactically integrated into the subordinate clause, there is no restriction against SS imposed by Particle, as in the \textit{wa} conditions. As such, it seems as if the increased looks to the TS referent found in all conditions at CONJ (including a slight rise in \textit{waNull}) endure longer in the
gaNull than in the waNull, even when participants ultimately choose SS, although looks to TS in WP gaNull are not as high as in the Overt conditions.

Again, while it is difficult to perform inferential statistics on the data by With/Against Prediction due to the uneven number of trials, a tentative examination of the descriptives (means and confidence intervals) for each critical time region confirms the general trends suggested in the time-course graphs:

![Graphs](image1.png)

**Figure 64 & 65.** L1 Target advantage scores for trials answered against and with prediction for Pronoun (DO2) & (V2).

Means for AP waOvert reflect the fact that participants fixate on TS by the end of the sentence, with looks to TS sharply increasing in V2 compared to DO2. Only gaNull remains noticeably higher in its AP mean (left panel) than its WP mean (right panel) by V2. Overall, the means confirm the descriptive observations above; by V2 all conditions but gaNull look very similar in both With and Against Prediction trials. Furthermore, even with high TS for gaNull, discounting the anomalous waOvert in DO2, there is an overall pattern in which Overt trials have higher target advantage means than their Null counterparts regardless of answer. In other words, ET data during the test sentence suggests that participants process the Pronoun according to
prediction as they hear it, regardless of their final interpretation decision, with the exception of \textit{gaNull} (where the WP and AP seem to trend with ultimate choice).

Results for the Test Sentence suggest that in at least three conditions (\textit{waOvert}, \textit{gaOvert}, \textit{waNull}), participants decide on their ultimate choice later, perhaps during the Question. However, this is not quite the case. As recounted in the WP/AP time-course graphs for the Question in Section 6.5.1, L1 participants seem to go into the Question fixating on their ultimate response. Indeed, collapsing the ET data over the Question reveals greater TS looks in the Null AP, and in the Overt WP. This suggests that participants “change their mind” at some point in between the Test Sentence and the Question in the Overt and in \textit{waNull}:

![Figure 66. L1 Target advantage scores for trials answered against and with prediction for Pronoun (Question).](image)

However, the separation between Overt and Null conditions is not nearly as clear during the Question in the Against Prediction trials as it is in the With Prediction trials. Furthermore, while \textit{gaNull} target advantage is higher in the AP trials than the WP trials, the preference is not clear cut, hovering around zero (meaning vacillation between TS and SS), much lower than \textit{waNull}. \textit{waOvert} also remains low, with looks hovering around zero as well. In the post-Question time
region, Overt and Null conditions do separate out a little further (\textit{gaOvert} even reaches a similar number as its With Prediction counterpart, \textit{gaNull}), \textit{gaNull} reaches the same target advantage level as \textit{waNull}, and \textit{waOvert} drops below zero.

![Graph](image)

**Figure 67. L1 Target advantage scores for trials answered against and with prediction for Pronoun (AQ).**

Overall, the L1 data divided into WP and AP trials reveals various amounts of processing difficulty in the different conditions. First, WP \textit{waNull} shows very little processing difficulty. Although there is a brief rise in TS looks after the CONJ, TS looks quickly dip again, and SS looks predominate throughout the Question and AQ periods as well. AP \textit{waNull} has very few trials, and the processing data also seems to reveal little difficulty: a clear preference for SS during the Test Sentence but then a clear preference for TS during the Question, suggesting that these trials represent some kind of error (such as misremembering which referent was mentioned in the Test Sentence). In the WP \textit{gaNull}, participants vacillate slightly during the matrix clause but clearly fixate on SS in the Question and After Question periods. Next, the WP and AP Overt trials show similar amounts of processing difficulty, with the locus of vacillation differing slightly between the two during the Q/AQ periods (longish vacillation during the Question for
AP Overt; slight vacillation at RT for WP Overt). The AP *waOvert* condition also seems to cause extra processing effort compared to AP *gaOvert*, most likely an effect of the interaction with Particle. Finally, the AP *gaNull* shows the most processing difficulty, with vacillation during the Question and extreme vacillation in the AQ period, suggesting that participants are consciously considering which answer to select.

6.6.3 *Eye-tracking Data by WP/AP Trials: L2 Group*

For L2 participants, unlike their L1 counterparts, WP trials and AP trials show opposite patterns as early as DO2 and continuing through V2 in all conditions.

**Figures 68 & 69. L2 Target advantage scores for trials answered against and with prediction for Pronoun (DO2) & (V2).**

When answering With Prediction (96% of Null; 19% of Overt), L2 participants have higher target advantage in the Overt compared to the Null; when answering Against Prediction (4% of Null, 81% of Overt), they show higher target advantage in the Null than Overt. In other words, L2 patterns during the Test Sentence, unlike those for L1 participants, seem to reflect their ultimate choice. Another way to say this, is that when L2 participants do answer With Prediction, their ET data during the Test Sentence tends to resembles that of L1 participants more than it
does when they Against Prediction (although the pattern is not as clear cut in the Overt conditions as it is in the L1 data). This suggests that when L2 participants *do* answer like L1 participants, they are also processing like L1 participants.

The time-course graphs for L2 *gaOvert* seem to back up this observation, with looks to the TS referent overtaking those to the SS referent in the matrix clause, although *waOvert* is a little more troublesome, as it was in the Q/AQ periods, reflecting the Particle issue.

![Graph](image)

**Figure 70.** Eye-tracking data for Test Sentence (*gaOvert*-With Prediction, L2).

![Graph](image)

**Figure 71.** Eye-tracking data for Test Sentence (*gaOvert*-Against Prediction, L2).

(L2 Null graphs are excluded as the low amounts of data makes them difficult to read; I have argued elsewhere that L2 AP Null largely represents error).
The L2 group is not quite the same as the L1 group during the Question, with the biggest difference being processing difficulty in the WP waOvert vs. difficulty in AP waOvert for L1 participants. Examining AP/ WP processing differences between groups in the Null doesn’t add much insight beyond the interpretation data, in that the L2 participants rarely go against prediction in the Null in the first place (note the huge variability in the AP gaNull during Q, further supporting the idea that the AP Null trials for L2 participants represent error).
Figures 74 & 75. L2 Target advantage scores for trials answered against and with prediction for Pronoun (Question) & (AQ).

In sum, separating the data into With Prediction and Against Prediction trials suggests:

- **AP gaNull** causes the most processing difficulty for L1 participants out of all other combinations, including increased reaction times and more vacillation during the Question and After Question periods. Choosing With or Against Prediction in the Overt seems to have less effect on processing cost, although there may be an interaction with Particle in the AP *waOvert* that creates some extra difficulty.

- L2 participants show most processing cost when answering topic-shift in the Overt conditions, as evidenced by longer reaction times in both *gaOvert* and *waOvert* compared to the Null, and more ambiguous looking patterns during and after the Question in the WP *waOvert*.

- ET data during WP *gaOvert* suggests that L2 participants process the pronoun similarly to L1 participants, with differences in *waOvert* arising from issues with the Particle, not the Pronoun.
6.7 Summary of Results

6.7.1 Pronoun

The L1 group followed prediction for the Pronoun condition, with an overwhelming choice of topic-shift (TS in 76% of trials) in the Overt and same-subject in the Null conditions (TS in 24% of trials). The interpretation data was backed up by the processing data, as L1 participants showed increased target advantage in the Overt conditions compared to the Nulls during the critical clause (DO2 and V2), and during the Question and post-Question periods. Taken with the results of previous studies cited in Chapter 4, this supports the PCI analysis, since a context designed to allow the calculation of the implicature as topic-shift worked, whereas ambiguous contexts in previous studies did not.

6.7.2 Particle

In the interpretation data, L1 participants followed prediction for Particle in the Null, with a higher TS rate in gaNull than waNull but not in the Overt, as gaOvert and waOvert had almost identical TS rates. However, choosing TS in the gaNull condition may entail greater processing cost, as gaNull trials in which L1 participants ultimately chose TS rather than SS (as predicted by pronoun) had very long reaction times that were accompanied by a period of vacillation between looks to TS and SS in the AQ period, suggesting that participants may have been second-guessing (perhaps consciously) their decision to choose a TS referent. This processing cost suggests that overriding the GCI (which biases SS) causes heavy processing costs. Meanwhile, the L1 ET data also revealed some effects of Particle in Overt condition, despite the lack of any effect in the interpretation or RT data. Specifically, there seems to be some increased processing difficulty in AP waOvert, reflecting the fact that the prediction for Pronoun in this condition (TS) is at odds with the syntactic constraint inflicted by the Particle
condition (SS). However, it is interesting to note that this processing cost from going against the syntactic constraint imposed by the topic particle is not as severe as that incurred by letting context override the GCI in the Null. That is, a syntactically clunky sentence incurs less processing cost than a non-typical interpretation for a grammatically-sound sentence.

6.7.3 Group

Results for Group largely followed prediction in that L2 participants did not show an overwhelming (> 50%) preference for topic-shift in the Overt condition, with a TS rate of 19% (there was no effect of Particle in the L2 interpretation data). L2 participants did choose TS in the Overt at a greater rate than they did in the Null (4%), a trend that was backed up in the eye-tracking data, with identical patterns to the L1 group during the Pronoun itself (PR), and with more TS looks in Overt than Null conditions in DO2 and V2, as well as during the Question. However, L2 TS looks were consistently lower than those of L1 participants in all regions (except PR), and target advantage was always negative, regardless of condition. In keeping with their general reluctance to do topic-shift, L2 participants, unlike L1 participants, are rather single-minded in Null condition, almost always choosing a SS referent (significantly more so than L1 participants in both Null conditions), and with very fast reaction times; the only significant reaction-time difference in all of the data was Null vs. Overt in the L2 group. L2 participants were also reluctant to violate the syntactic constraint imposed by the topic-marker in the waOvert condition, as reflected in the fact that WP waOvert showed much more vacillation than WP gaOvert, which patterned much more like the L1 data.

Three L2 outliers had interpretation data for the Overt condition similar to the L1 mean, although they still choose TS very infrequently in the Null. It is difficult to analyze what exactly set these three participants apart, although proficiency is most likely not a factor. One of these
three participants was tested in Japan, suggesting that current input may be an important factor, and that a further study testing L2 participants in Japan vs. in the United States may be fruitful.
CHAPTER 7
DISCUSSION

7.1 L1 Group

7.1.1 L1 Interpretation Data

The interpretation data revealed that L1 Japanese participants consistently chose a topic-shift interpretation for overt pronouns (76% of the time) in this experiment. These results stand in contrast to previous experiments that showed no antecedent preference for 3rd person pronouns for native Japanese speakers, with participants choosing topic-shift or same-subject interpretations at chance (Kanno, 1997; Okuma, 2012; Ueno & Kehler, 2010). Taken together, the results of this study and these previous studies support my hypothesis that the inclusion of a controlled context that strongly supports topic-shift is necessary to consistently elicit a topic-shift interpretation for overt pronouns in Japanese, in line with the PCI analysis outlined in Chapter 4. Furthermore, since this specialized context is necessary to achieve a more-often-than-not topic-shift rate in Japanese but not in Italian, the interpretation of overt 3rd person pronouns in the two languages differs; I argue that this difference can be best modeled via the distinction between particularized (Japanese) and generalized (Italian) implicatures. Review of the processing data in the next section will add support to this analysis.

That the context provided in this experiment is effective in encouraging a topic-shift interpretation is also evident in the Null condition, since this context elicited a topic-shift reading 36% of the time in the gaNull, despite the fact that Nulls bias a same-subject interpretation. In a GCI-style anaphora system and other types of anaphora hierarchies, nulls (especially nulls without accompanying agreement morphology) are the most extremely minimal forms (phonetically, morphologically, semantically), and thus are biased to retrieve highly
local/accessible/salient antecedents. However, as conversational implicatures are cancellable, under the GCI analysis, speakers can use context (in this experiment, the presence of the two characters presented both linguistically and visually) to cancel the same-subject bias for Nulls, with the processing data suggesting that doing so carries great processing cost (as if participants are overthinking their response).

The secondary Particle condition holds in the Null but not the Overt. Recall that wa-marked (=topic-marked) nominals cannot serve as the subject of a subordinate clause, and must instead scope over the entire sentence. As such, it seems that the topic-marked NPs in the waNull condition blocked the topic-shift interpretation: since the topic is established in-sentence by the wa-marked NP, without a second NP in subject position, participants had little incentive to shift to something else, regardless of context. Interestingly, predictions for Particle were not borne out in the Overt interpretation data, with roughly equal proportions of topic-shift in waOvert (76%) and gaOvert (76%). In other words, the prediction for Pronoun (perhaps aided by the context biasing topic-shift) trumps the prediction for Particle for L1 participants in the waOvert.\textsuperscript{52} Note that in our system, this is an example of a pragmatic calculation (the PCI generated by the Pronoun) overcoming a syntactic constraint (on the scope of the wa-marked NP).

7.1.2 L1 Processing Data

Overall, the L1 processing data reflects the trends found in the interpretation data, with more looks to the TS referent in Overt than Null conditions in all time periods (DO2, V2, Q, AQ), as well as more TS looks in ga than wa in the DO2, the Q, and for gaNull and waNull the AQ periods. In further analyzing the processing data, it is important to note that the ET data

\textsuperscript{52} It is also possible that participants (implicitly or otherwise) re-interpreted the waOvert construction as the intrinsically contrastive –wa/-wa construction rather than a –wa/-ga construction, although I do not explore that possibility here.
bears out the effectiveness of the topic-shift context given in each test item. For L1 participants, there is an increase in looks to the TS referent (and/or a decrease in looks to the SS referent) during the conjunction and continuing into DO2, which occurs in all conditions, including waNull, which has the lowest TS interpretation rate (although the effect is most attenuated there), and regardless of ultimate interpretation (whether participants ultimately choose TS or SS). It seems as if the presence of the two possible referents (in both the introduction sentence and on the screen) coupled with the clause-ending cue introduced by the conjunction, is enough to lead participants to consider the relevance of the non-subject referent, which is then “boosted” by the overt pronoun when it appears. In the waNull condition (regardless of ultimate choice), TS looks quickly decrease at the onset of DO2; in gaNull With Prediction trials (ultimate SS choice), TS looks linger, but SS looks clearly dominate. In gaNull Against Prediction trials (ultimate TS choice) and both Overt conditions (regardless of ultimate answer), TS looks remain elevated and dominate over SS looks.

Next, breaking down the data by participants’ ultimate choice reveals different processing costs for breaking the Null bias vs. going against prediction in the Overt, which I suggest reflects the difference between the GCI-driven Null interpretation and PCI-driven Overt interpretation. Specifically, there is greater disparity in the ET data between With Prediction (WP) and Against Prediction (AP) trials in the Null condition compared to that between the WP and AP trials in the Overt. In WP Null trials (76% of trials), L1 participants fixate preferentially on SS throughout the Test Sentence; they then carry this interpretation over into the Question, with more looks to SS than TS from Question onset until (a very quick) mean RT. That is, they show very little doubts about their interpretation. On the other hand, fixation and reaction-time data reveal great
processing difficulty in the AP gaNull trials\(^{53}\) (36% of trials). In the AP gaNull, L1 participants look preferentially at the TS referent during the Test Sentence, in line with their ultimate interpretation. Participants then maintain this fixation preference for the TS referent from the Test Sentence into the Question; however, the ET data during the Question and After Question periods reveal a great deal of processing struggle (in line with the fact that AP gaNull had the longest reaction times of any condition; about 800 ms longer than the next slowest, AP waOvert). Specifically, during the Question itself, there is some vacillation between the SS and TS referents, although the target advantage score just barely remains positive when averaged over the entire Question (.09), and TS looks continue to dominate well into the After Question period. However, about 1250 ms after Question offset, participants show an extended period of vacillation between the TS and SS referents, lasting for almost 1000 ms, followed by a period where looks to the SS referent dominate, before participants finally choose their answer, at a mean of 3250 ms after Question offset.

![Figure 76. Eye-tracking data for Question (gaNull-Against Prediction, L1).](image)

\(^{53}\) As mentioned in the previous chapter, the few number of trials for AP waNull makes the fixation data difficult to analyze, but results suggest that these responses may represent error (not paying attention, etc.). Because this condition is syntactically constrained against TS, I largely set it aside in the discussion here, instead contrasting gaNull with the Overt conditions.
In other words, when L1 participants choose Against Prediction in the Null, they come to the TS interpretation readily during the Test Sentence, vacillate but more or less maintain the TS interpretation into the Question and even after the Question, but then (perhaps consciously) doubt this interpretation when it comes time to actively choose an answer. Overall, L1 participants show little processing difficulty during the WP (SS) Null trials but a great deal of processing difficulty during the AP (TS) gaNull trials (and they can’t even choose AP in the waNull, with topic-particle further discouraging the TS interpretation, although the topic-particle does not stop them from choosing TS in waOvert).

On the other hand, the amount of processing difficulty between the WP and AP trials in the Overt conditions seems roughly equal (although perhaps a little easier in the WP trials), and somewhere in-between the two Null extremes. In both WP (ultimate TS choice) and AP (ultimate SS choice) Overt trials, L1 participants begin to show increased looks to the TS referent during the conjunction, which continue to rise sharply during the Pronoun, and then dominate over SS beginning at DO2 and continuing through V2. That is, by the end of the critical clause, regardless of their ultimate answer, L1 participants fixate preferentially to the TS
referent in the Overt conditions. The WP and AP trials differ slightly during the Question. In the WP trials (76% of trials), participants continue to fixate on the TS referent from Q onset to offset; however, they seem to do a quick “double-check” around mean RT, with TS and SS looks very briefly leveling out during *waOvert* and almost (but not quite) leveling in *gaOvert*. However, this “double-check” is nowhere near as extreme as the vacillation in the AP *gaNull* trials, with no vacillation during the Question, slight vacillation at mean RT (not before it), and less extreme reaction times. Note the very clear TS fixation preferences from Question onset until mean RT:

![Figure 78. Eye-tracking data for Question (gaOvert-With Prediction, L1).](image)

![Figure 79. Eye-tracking data for Question (waOvert-With Prediction, L1).](image)
In AP Overt trials (24% of trials), participants show more vacillation during the Question itself, but preferentially fixate on SS by the end of the Question with no further uncertainty. (In gaOvert this manifests as TS and SS looks evening out; in waOvert TS and SS looks cross over at V2 onset).

![Figure 80. Eye-tracking data for Question (gaOvert-Against Prediction, L1).](image)

![Figure 81. Eye-tracking data for Question (waOvert-Against Prediction, L1).](image)

In other words, in WP trials, participants keep the interpretation they ended up with during the Test Sentence, but seem a little surprised by it at the moment when they make their decision. In the AP trials, participants change their mind from the interpretation reached in the Test Sentence.
to a different interpretation *during* the Question, and then remain satisfied with this choice until mean RT (with slightly more vacillation in *waOvert* due to the Particle).

It seems, then, that following prediction for Nulls is very easy, but going against prediction is very hard. On the other hand, answering with prediction in the Overt has slight processing cost at mean RT, while answering against prediction in the Overt has slight processing cost during the Question. These results can be interpreted to support the GCI/PCI distinction between Nulls and Overts hypothesized in Chapter 4. The GCI is more closely tied to structure, acting at the level between semantic (truth-conditional) and utterance-token meaning, so breaking it should be more difficult than the PCI, at the level of utterance-token meaning. In the Null, when participants ultimately choose according to the default same-subject interpretation, they show processing ease, but when they ultimately choose against prediction, they show severe cost, suggesting that cancelling the GCI is difficult. The Overt, on the other hand, shows minimal processing cost regardless of interpretation choice. In contrast to the WP Null trials, which show a clear preference for SS throughout the Question and After Question periods, even *With* Prediction Overt trials show some vacillation during the After Question period, suggesting that they are double-checking their choice. This “double-check”, however, is not as dramatic as the vacillation found in the AP Null. In the AP Overt trials, participants go into the question with a TS interpretation, but then seem to change their minds during the Question. However, cancelling the PCI here seems to take less effort than going against prediction (i.e., cancelling the GCI) in the Null. Note also that breaking the syntactic condition on the Particle in WP *waOvert* does show processing cost, but less so than breaking the GCI in the *gaNull*. 
However, further studies are necessary to definitively support the GCI/PCI analysis over other analyses in which interpretation of the Null is more constrained (either by syntax or by other pragmatic factors) than the Overt. Additionally, the exact timing of the calculation and/or cancellation of these implicatures remains unclear. Under my analysis of the data, the GCI is calculated and cancelled during the test sentence, but the PCI isn’t cancelled until the Question; in other words, the process of cancelling the GCI happens faster than the process of cancelling the PCI. Since the GCI is claimed to be a default, this timing could be possible, since the calculation of the implicature in the first place happens faster. It could also be the case that cancelling a GCI is a more complex process than cancelling a PCI, and therefore is longer in duration; in other words, the entire duration from critical clause through final interpretation could reflect the “cancelling” process for GCIs, whereas for PCIs it happens only during the formulation of final interpretation (i.e., during the Question). While neither of these hypotheses is out of the realm of possibility, (perhaps many) further experiments are necessary to pin down these processes.

Furthermore, there is no instance where the PCI fails to be calculated during the test sentence, even in the AP Overt trials, most likely because the context in this experiment is strong enough to elicit a topic-shift interpretation from the very start. Further experiments using the same test sentences in various contexts (no context, context that more definitively bias topic-shift using the actions in the sentence, contexts biasing SS interpretation) would be fruitful to further tease apart these issues. It would be particularly fruitful to see if giving participants my test

54 It also not out of the realm of possibility that participants are being primed by the many Overt pronouns appearing in the other trials, including fillers; the amount of 3rd-person pronouns they hear in this experiment is almost certainly greater than the amount they hear in the natural language input.
sentences without context in a paper-and-pencil task would result in a chance interpretation for overt pronouns, as predicted by the PCI analysis (and found in previous studies, although I didn’t directly test it in this study). Another follow-up study would examine whether changing the introductory sentence by leaving out the extra referent would lead to increased SS looks across the board (although there is no way to make this particular comparison without having both referents on the computer screen, thereby adding some context that supports topic-shift). Finally, without testing similar items in other languages using a similar experiment, there is no way to directly compare the PCI-style overts in Japanese to GCI-style overts in other languages.

7.2 L2 Group

7.2.1 L2 Interpretation Data

Unlike the L1 Japanese participants, L1-English/L2-Japanese participants did not show an overall tendency to topic-shift in the Overt condition (TS interpretations 19% of the time). In fact, L2 participants showed a general resistance to topic-shift, as their rates in Null were low as well (4%). These results, overall, are in line with previous studies done on the Interface Hypothesis. However, there are two differences in the results of Overts: (i) the disparity between rate of topic-shift in the overt by L1 and L2 participants is greater in this experiment (20% difference in Sorace & Filiaci (2006); 25% in Belletti et al. (2007); 57% in this study), reflecting the fact that L2 participants in this study did not choose a TS interpretation for overts more often than not (19% in this study vs. 73% and 79% in previous studies), and (ii) three L2 outliers in this experiment show native-like behavior in the overt, whereas the previous studies report no
such outliers.$^{55}$

Table 16. Topic-shift rates found in results of Tsimpli, et al., 2004, Sorace & Filiaci, 2006, Belletti, et al., 2007, and This Study

<table>
<thead>
<tr>
<th></th>
<th>Topic-shift rates</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Overt</td>
<td>Null</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$L1$/Monolingual</td>
<td>$L2$/Bilingual</td>
<td>$L1$/Monolingual</td>
<td>$L2$/Bilingual</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(wa/ga)</td>
<td>(wa/ga)</td>
<td>(wa/ga)</td>
<td>(wa/ga)</td>
<td></td>
</tr>
<tr>
<td>Tsimpli et al. (2004)</td>
<td>92%</td>
<td>79%</td>
<td>49%</td>
<td>30%</td>
<td></td>
</tr>
<tr>
<td>$L1$-Italian/$L2$-English bilinguals</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sorace &amp; Filiaci (2006)</td>
<td>92%</td>
<td>73%</td>
<td>49%</td>
<td>54%</td>
<td></td>
</tr>
<tr>
<td>Near-native $L2$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Belletti et al. (2007)</td>
<td>95%</td>
<td>70%</td>
<td>60%</td>
<td>46%</td>
<td></td>
</tr>
<tr>
<td>Near-native $L2$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>This Study</td>
<td>76% (76%/76%)</td>
<td>19% (19%/19%)</td>
<td>24% (36%/12%)</td>
<td>4% (4%/3%)</td>
<td></td>
</tr>
<tr>
<td>Advanced $L2$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In other words, unlike in the other IH studies, L2 participants do not seem to be displaying residual optionality in their interpretation of overt pronouns; their interpretations are either completely different form L1 participants, or exactly the same.

It is important to note that the differences in this study compared to previous studies could have arisen from the different structures used in the test sentences. In the former studies, participants shifted from the subject to a non-subject antecedent from the previous clause; in this study, they shifted to an external referent. However, it is also possible that the difference between the Japanese L2 and Italian L2 groups could arise from different constraints on overt pronouns in the two languages. Under the analysis put forth in this study, L2 Japanese speakers

$^{55}$ Recall Rothman (2009) also found native-like overt pronoun use by L2 speakers of Spanish, using different tasks. His results were generalizable to the entire advanced L2 group, however, not just an outlying subset of participants.
have a different task in learning how to use overt pronouns compared to L2 Italian speakers; instead of “retuning” the setting for the overt pronoun for their L2 anaphora scale, they must learn to take the overt pronoun off the anaphora scale completely. We will return to this idea in Section 7.3.2.

In addition to the differences in the Overt condition, L1 and L2 groups in this study (including the three native-like L2 participants) also diverged from each other in Null condition, as opposed to the results from Sorace & Filiaci (2006) & Belletti et al. (2007), in which L1 and L2 groups did not differ in the Null. In this study, L1 participants were willing to override the null default approximately 36% of the time in the gaNull condition, while the L2 participants only did so about 4% of the time. Note that Tsimpli et al. (2004) did find that their bilingual group chose less topic-shift interpretations in the Null compared to monolinguals. However, Tsimpli et al. (2004) suggest that the differences in the Null stem from bilinguals reanalyzing the critical clause as non-finite (i.e. *While crossing the street...* instead of *While [pro] crosses the street*) and subsequently treating it like a non-finite clause in English, despite the presence of finite verb morphology. This analysis doesn’t quite work for my materials, since the Null/Overt alternation is in the matrix clause, not the subordinate clause. I do not postulate different underlying mechanisms for Null pronoun interpretation in Japanese and Italian, treating both with a GCI analysis. However, as noted in previous chapters, the rates of null use in Japanese is much higher than in other languages, and, as noted in Ariel (1990), *pro* with agreement morphology (as in Italian) has more structure (and therefore is lower on an anaphora scale) than nulls with no agreement morphology as in Japanese. These differences are reflected in the fact
that even L1 Japanese participants give topic-shift interpretations much less frequently than in the Italian studies.\footnote{Levinson (2000) theorizes that GCIs may represent a point in the grammaticalization process (between pragmatics and syntax). Under this analysis, we could argue that the null bias for same-subject in Japanese has become grammaticalized (or is closer to being grammaticalized) than that in Italian. In this case, the AP Null trials represent not the cancelling of a GCI, but context overriding a syntactic constraint. While this idea may be worth examining in a future study, I do not explore it further here.}

7.2.2 L2 Processing Data

The L2 processing data sheds further light on the interpretation differences between L1 and L2 groups. First, the fixation data suggests that L2 participants are less sensitive to context than the L1 group, especially in the Null condition. While the TS-biasing context increased looks to the TS referent in all conditions of the L1 data, including the Null conditions, regardless of ultimate interpretation choice, this effect was not found in the L2 data. While there were increased looks to the TS referent in both Overt conditions, there was no increase in TS looks in the Null conditions, with L2 participants unequivocally fixating on the SS referent in the Nulls. This suggests that the L2 participants were overall less sensitive to contextual factors than the L1 participants, who had increased TS looks during CONJ and DO2 even in waNull and AP gaNull. Indeed, the processing data reflects the fact that unlike L1 participants, L2 participants very rarely choose topic-shift in the Null, even in the gaNull. In both waNull and gaNull, L2 participants fixate on the SS referent; there are peaks in TS looks during DO2, but they are much lower than their counterparts in the L1 data.\footnote{The difference between wa.L2 and wa.L1 is actually not significant during DO2, but as there is no three-way interaction, it is difficult to tell if this lack of significance is due to waOvert rather than waNull. L2 waNull has a raw target advantage score of -.22 and L1 waNull -.15.} Unsurprisingly, L2 target advantage scores in the Null condition are very low throughout the Question (-.37 ga and -.34 wa) and After Question periods (-.7 ga and -.56 wa), and the Null conditions have very low reaction times. In other
words, while L1 participants let context influence (and sometimes override) the same-subject bias for the Null as early as the Test Sentence, incurring processing cost when they ultimately choose against prediction, L2 participants are less likely to do so. Furthermore, the WP waOvert, in which context trumps the syntactic constraint imposed by the Particle, L2 participants show extra processing cost (whereas L1 participants show the opposite—more processing cost when the Particle trumps the context-induced TS bias of the Pronoun).

While L2 participants do not show an overwhelming TS interpretation preference for the Overt conditions, their fixation data makes it clear that they did not process Overts and Nulls in the same way. The fact that the L2 group chose TS more often in the Overt (19%) than Null (4%) is reflected in the eye-tracking data, with greater target advantage in the Overt conditions compared to the Null for both L1 and L2 groups during DO2, V2, and during the Question. In the Null conditions, looks to the SS referent are much higher than they are in the Overt, and looks to the TS referent much lower. In the Overt conditions, on the other hand, looks to the TS referent increase during CONJ and then continue to rise during the Pronoun, just as they do in the L1 data, even in AP trials where L2 participants ultimately choose the SS referent:

58 In the AQ (After Question) period, this trend is somewhat complicated; both groups have higher target advantage in the Overt compared to Null for the ga conditions, but only L1 participants do so in the wa conditions. The lack of a clear TS fixation preference in waOvert during the AQ period most likely reflects the influence of Particle discussed below.
Indeed, in the Overt conditions, there are no significant differences between L1 and L2 groups in target advantage during the Pronoun itself. However, in the L2 but not L1 data, TS and SS looks begin to even out during DO2, with L2 participants clearly fixating more on SS than TS for the entirety of V2. L2 participants carry this preference for the SS referent into the Question, with looks to SS dominating throughout, into the After Question period, and at mean RT. In other words, even in the AP Overt trials (ultimate SS choice), L2 participants notice the overt
pronoun in the same way that L1 participants do, but they react to it differently. This result would be in keeping with the PCI/GCI analysis given in this paper. The L1 participants, upon hearing the Pronoun, take note of its presence and deal with it by calculating an implicature for topic-shift that then gets confirmed or canceled in the Question period. On the other hand, the L2 participants in AP Overt trials, upon hearing the Pronoun, take note of its presence and deal with it by falling back on the English anaphora scale, where an overt pronoun is available. We explore this further in section 7.3.2 below.

When L2 participants do answer With Prediction in the Overt (ultimate TS answer), they look more like L1 participants, but show more variation across Particle. In WP gaOvert, unlike AP Overt trials, TS looks not only increase but also overtake SS looks during the Pronoun, and, despite some fluctuation in DO2, end up dominating over SS looks throughout V2. During the Question, they initially continue to fixate preferentially on the TS referent, but then seem to reconsider the SS referent during V2, before ultimately fixating on the TS referent again in the AQ period. Unlike with L1 participants, there is little evidence of a “double-check,” and reaction times are 400 ms longer than in L1 WP gaOvert. In other words, L1 and L2 WP gaOvert trials show similar patterns, especially during the Test Sentence, but L2 participants seem to reconsider and confirm their interpretation during the Question, whereas L1 participants may quickly confirm their answer while clicking at mean RT (without SS overtaking TS as in the L2 data during Q). Overall, the L2 WP gaOvert fixation data is quite similar to its L1 counterpart, albeit with slightly more processing difficulty during the Question (including a longer reaction time). This suggests that when L2 participants do ultimately choose a topic-shift interpretation in gaOvert, they also process the overt pronoun like L1 participants, perhaps with a bit of extra effort.
Figure 84. Eye-tracking data for Test Sentence (gaOvert-With Prediction, L2).

Figure 85. Eye-tracking data for Question (gaOvert-With Prediction, L2).

The WP waOvert trials, on the other hand, show more differences between groups in the processing data. Specifically, L2 fixation patterns during the Test Sentence do not look very different in the WP and AP in waOvert; indeed, during V2, the average target advantage score is more or less the same in both WP and AP trials (.23). Furthermore, looks patterns during the Question and After Question periods reveal a great deal of vacillation throughout, as if
participants are consciously considering their answer. While the target advantage scores for Q and AQ are positive, they are close to zero (.03 Q and .05 AQ).

![Figure 86. Eye-tracking data for Test Sentence (waOvert-With Prediction, L2).](image)

![Figure 87. Eye-tracking data for Question (waOvert-With Prediction, L2).](image)

While choosing against prediction in the *gaNull* was most difficult for L1 participants, choosing with prediction in the *waOvert* seems to be most difficult for L2 participants. Although L2 participants ultimate choose TS at the same rate in *gaOvert* and *waOvert*, it seems as if the Particle constraint encouraging a same-subject response exerts extra processing effects in the latter when choosing a TS interpretation. That is, L2 participants are less sensitive to context
when doing so means going against the morphosyntactic constraints on Particle. In other words, when L2 participants do choose TS for the Overt, their data is similar to that of L1 participants, but with this similarity modulated by Particle effects in the *waOvert*.

### 7.3 Explaining L1/L2 Divergence

#### 7.3.1 Nulls

L2 participants interpret null pronouns according to prediction, and are more reluctant to go against this prediction than L1 participants. The biggest question in the Null data is why the L1 group was more willing to override the default interpretation more often than the L2 group, including even the L2 participants who resembled L1 participants in the Overt. There are several possibilities, but I argue that it is the act of overriding the null bias (GCI) using context that is problematic for L2 participants. We first explore the alternative explanations. First, L1/L2 divergence in the Null may be related not to the Pronoun condition but to the Particle condition. L1 participants chose a topic-shift interpretation for the Null by far more often in *gaNull* (36%) than *waNull* (11%). If L2 participants treat both *ga* and *wa* the same as *wa*, it could explain the lower topic-shift rates. This explanation would suggest that L2 data in both Null conditions should resemble the L1 *waNull*, which to some extent it does. There are no significant differences in target advantage score between L2 *wa*/L2 *ga* and L1 *wa* in DO2, or any significant differences between L2 *waNull*/L2 *gaNull* and L1 *waNull* in V2. However, the interpretation data shows that L2 participants choose TS significantly less in both Null conditions (4% *wa*, 3% *ga*) than in L1 *waNull* (11%). Furthermore, it is clear that L2 participants do process *wa* and *ga* differently in the Overt condition, as described in the previous section, suggesting that they do not uniformly treat the two particles as the same.
The next possibility is that L2 participants have a consistent strategy for the same-subject referent regardless of Pronoun type, under the theory that maintaining reference is less costly than switching reference. However, in this case, those Overt trials in which participants chose a same-subject referent (AP Overt) should resemble the Null trials, which it does not: during DO2 the target advantage score for WP Nulls is -.26, compared to -.06 in the AP Overt. If such a uniform processing strategy is in place, it is saving participants from processing struggle more in the Null than the Overt for some reason.

Third, L2 interpretation and processing data in the Null condition could reflect transfer from English. Although null pronouns are rare in English, they do exist in limited circumstances (non-finite clauses, conjoined clauses such as Jane went to the park and ate lunch) and, like their Japanese counterparts, bias a same-subject reading. However, nulls are not found in the construction used in my test sentences (While Takeshi drinks a beer, looks at the scenery is not a grammatical English sentence). Furthermore, a transfer account alone doesn’t explain why three L2 participants were able to free themselves from the influence of English for the Overt but not the Null.

Finally, L1/L2 divergence might be rooted in the process of using context to override the Null default (GCI) interpretation. As argued in Section 7.2.2, L2 participants seem less sensitive to context than L1 participants even during the Test Sentence. It is also clear from the L1 data that choosing against prediction in the Nulls carries serious processing cost, with participants perhaps consciously reconsidering their choice before selecting their final answer—almost that they are overthinking their response. L2 participants, on the other hand, know a “rule” for Nulls and don’t go looking for reasons to break it. The divergence between L1 and L2 groups in the Null, therefore, seems to stem from the reluctance of the latter but not the former to go against
the same-subject bias in the absence of a reason stronger than the context provided in the test sentence.\(^{59}\) This could be considered a processing strategy ("don’t go looking for trouble"), but it is specifically about following the GCI (default) for the Nulls, not a general same-subject processing strategy for all anaphors. (In the Overt condition, on the other hand, there is no consistent rule to follow, so there is nothing to break).

### 7.3.2 Overts

L1 and L2 participants clearly diverge in their interpretation of Overt pronouns, with the former overwhelmingly choosing a topic-shift interpretation and the latter overwhelmingly choosing a same-subject interpretation. Like in the Null condition, there are several possible explanations: processing, transfer, and difficulty using context to calculate the PCI. As argued in the Nulls, it does not seem to be the case that L2 participants are falling back on a general maintain-reference processing strategy, since L2 participants do not process the Overt in the same way as the Null, even when ultimately choosing the same-subject interpretation. Even in the AP Overt trials (ultimate SS interpretation), there is an increase in TS looks during the Pronoun and DO2, with the drop-off not occurring until V2, whereas in the Null TS looks remain low throughout the Test Sentence. Furthermore, reaction times in gaOvert are similar in WP and AP Overt trials (1963 vs. 1803 ms), even though under a processing account the former should be more difficult to process than the latter. If the same-subject strategy is supposed to be saving processing effort, it doesn’t seem to be doing much, at least in the gaOvert. Choosing TS does cause processing difficulty in the waOvert, where there is a conflict between Pronoun (TS) and Particle (SS). There is some vacillation in the ET data after the Question in waOvert, with

---

\(^{59}\) Furthermore, if the L1 group arrives at TS interpretation for the Null via conscious effort, the lack of this effect in the L2 group may be tied to the L2 participants’ lack of confidence in overriding a rule that is very rarely violated in natural conversation.
divergence in reaction times between AP (1730 ms) and WP (2962 ms) trials. However, as argued above, this is due to the interaction of Pronoun with Particle, rather than the pronoun itself. A general same-subject processing strategy, therefore, seems unlikely.

This leaves transfer and failing to calculate the PCI as possible accounts; I argue that the best account is a combination of both. If we follow the PCI/GCI analysis I have put forth, L1/L2 divergence is rooted in the respective reaction of each group to the Overt pronoun. L2 participants do notice the presence of the Overt pronoun, and react to it, as indicated by looks to TS that increase during the Pronoun and DO2 before dropping sharply in V2. However, if L2 participants were simply failing to calculate the PCI, their interpretation would end up at chance (like Japanese participants in the previous studies cited in Chapter 4), not with a preference for the subject referent. Rather, the ET data suggests that L2 participants notice the Pronoun, consider it, and then interpret it as they would in English. L2 participants notice the overt pronoun, realize they have no place for it on their Japanese anaphora scale, and then fall back on their English interpretation in order to deal with the problem (a resolution perhaps encouraged by their general reluctance to let context influence their interpretation, as also seen in the lack of TS interpretations in the Null). L1 participants, on the other hand, notice the pronoun, realize they have no place for it on their Japanese anaphora scale, and react by calculating a PCI; they do not have the alternate route (English anaphora scale, with a very strongly tuned preference for overt pronouns) that L2 participants have. In other words, L2 participants display transfer, but it is a kind of “last-resort” in the absence of a clear interpretation pattern for the Japanese 3rd person overt pronoun, and in their reluctance to use context to influence their interpretation of pronominal forms.
The next question is the exact nature of this transfer effect, and whether or not L2 speakers can shed it. The data suggests that they can, since three outliers did show native-like Overt interpretation patterns. Sorace (2011) postulated both priming and antipriming as mechanisms affecting the mental representation of overt pronouns in bilinguals. This analysis might explain why L2 participants in this study, even highly proficient ones, have much lower TS rates than their counterparts in the Italian studies (-57% vs. -20-25%). The Italian bilinguals can rely on both priming of the Italian pronoun and antipriming of the English pronoun in order to arrive at a representation that more closely approximates the Italian representation, whereas Japanese L2 speakers can only rely on the antipriming of the English pronoun, since there isn’t really anything to prime in the Japanese, due to the fact that Japanese 3rd person pronouns do not have a default interpretation. In other words, while exposure to increased Japanese input may or may not be useful, it is critical that L2 Japanese speakers have decreased English exposure in order to suppress (weaken activation of) the English pronoun. Crucially, while the Italian speakers can rely on increased L2 input and/or decreased L1 input in order to tune themselves to Italian-like pronoun settings on the anaphora scale, L2 Japanese speakers must rely on decreased English input in order to arrive at a Japanese-like interpretation.

Under this analysis, this “retuning” may happen when the nature of the primary linguistic input changes, as has effectively been argued in previous IH studies (Tsimpli et al., 2004). For example, an L2 speaker who has been primarily been using Japanese for a year may show high levels of topic-shift, since their English anaphora scale is suppressed, and they therefore do not automatically fall back on it when encountering Japanese pronouns. Upon returning to an

---

60 This could very well describe the case of participant 709, tested in Japan, who was attending graduate school in a Japanese university. Other studies have also suggested individual cognitive
English-speaking environment, however, this same speaker might revert back to English-like interpretations. In other words, while the group differences in this study may result from a type of transfer, it is a variable one, highly sensitive to the speaker’s current linguistic environment. Note that this explanation can also help explain the differences in the Null and Overt condition: suppressing the English pronoun interpretation preference is a different process from letting context overriding the GCI in the Null. It is difficult to see how the latter process would be tied directly to current input (other than a general confidence that comes with increased proficiency in a language).

7.4 Summary of Experiment

The interpretation, reaction-time, and fixation data in this study (especially taken together with previous studies) support the particularized conversational implicature (PCI) analysis of anaphora in Japanese that I proposed in Chapter 4 of this paper. In this experiment, a context designed to support the calculation of a PCI that switches reference from the subject to another referent succeeded in eliciting a consistent (76%) topic-shift interpretation for overt pronouns from L1 Japanese participants, where previous experiments without such context found participants closer to chance. Furthermore, reaction-time and eye-tracking data reveal a more extreme effect of interpretation choice for null compared to overt pronouns. Topic-shift in the Null is very costly, while maintaining reference is easy, whereas in the Overt conditions, processing difficulty is somewhere in between for both choices. In other words, cancelling the GCI in the null is more difficult than calculating/canceling a one-off PCI in the overt. More experimentation using different contexts is needed to confirm for certain that this effect is best

---

differences in inhibition capacity may also lead to individual variation in L2 pronoun interpretation patterns, which could explain the other two outliers (Sorace, 2011).
captured by the PCI/GCI distinction (one remaining mystery is the timing of when participants calculate and cancel the GCI vs. the PCI). Running a similar experiment in a language with GCI-style overts would also allow for a useful comparison.

This study also revealed a wider gap between L1 and L2 groups than in previous studies on overt pronoun use in bilingual Italian populations. In particular, L2 participants tended to fall between two extremes: very low rates of topic-shift (19%) or rates comparable to L1 participants (three L2 participants with average TS rate of 77%). Unlike in the Italian studies, L2 participants do not seem to be showing increased optionality in overt pronoun interpretation; instead, they either look like L1 participants or they don’t. This result is also not out of line with a PCI analysis of Japanese overt 3rd person pronouns, as under this analysis Japanese and Italian learners have different tasks. L2 Italian learners must sort out where the overt pronoun fits on the anaphora scale in Italian relative the scale in English, adjusted by priming and antipriming effects from the input, with some residual overlap (perhaps exacerbated by processing pressures). L2 Japanese learners, however, do not need to adjust the position of the overt pronoun on the anaphora scale in Japanese, because it is not on the scale. Instead, in order to display native-like interpretation patterns, they must suppress an English-like reaction to the overt pronoun in the absence of a default interpretation for the Japanese pronoun; once this reaction is sufficiently suppressed, the Japanese-like interpretation emerges, as suggested by the three native-like outliers in this experiment. As such, the interpretation of overt pronouns by L2 Japanese speakers is largely constrained by their relative proportions of input to the L1 and the L2; in particular, by the deactivation of English due to decreased English exposure. As most of my participants were tested in the United States, it is therefore unsurprising that most of them did very little (or no) topic-shift (or that the only participant tested in Japan had Japanese-like interpretations). A future
study comparing L2 groups tested in Japan and in the United States would provide further evidence for this analysis.

7.5 Overall Summary

In this study, I argue that the distinction between particularized (PCIs) and generalized conversational implicatures (GCIs) can model the differences in overt 3rd person pronoun interpretation by native speakers of Italian and Japanese found in various experimental studies, as well as the differences among advanced L2 speakers of these two languages. Both types of implicature are based on the idea that using an overt pronoun where a null could be used instead violates Gricean Maxims (Grice, 1975). However, GCIs are default implicatures associated with the complement of the (maintain-reference) null interpretation, i.e. topic-shift. For PCIs there is no set implicature; the hearer must construct a novel implicature directly from context. As implicatures, GCIs can be canceled by context, but unlike PCIs, the implicature does not need to be calculated from scratch based on context.

In Italian (and some other European null-subject languages), overt 3rd person pronouns have been found to consistently elicit a topic-shift interpretation in experimental studies, whereas in experiments on Japanese, overt 3rd person pronouns have elicited ambiguous interpretation patterns, with participants split at chance between topic-shift and maintain-reference interpretations. Crucially, the topic-shift interpretation for overts in Italian is stronger in ambiguous contexts than unambiguous ones, which would be expected under a GCI analysis, as the default interpretation emerges in situations where the GCI is not overridden by other contextual (disambiguating) factors. In the Japanese experiments, on the other hand, participants are at chance in ambiguous contexts, which would be expected under the PCI analysis, since hearers are expected to generate an implicature from scratch with each utterance, with no default
interpretation to fall back on; the lack of context (or lack of a context that consistently supports one kind of implicature over others) leads to ambiguity. I also offered properties of Japanese overt 3rd person pronouns demonstrating how they differ from pronouns in other languages.

In my experiment, I created a context that consistently supports topic-shift by providing two (and only two) available referents in an introduction sentence. In the results, Japanese native speakers chose topic-shift for overt pronouns at a level greater than chance (76%), in contrast to previous studies with no such context. The eye-tracking data conforms to the patterns in the interpretation data, with more looks to the TS referent than the SS referent in Overt compared to Null conditions. The processing data further supports the distinction between the GCI implicature in the Null and the PCI in the Overt as breaking the former causes more processing pressure than either choosing or not choosing topic-shift in the Overt, although further experiments are needed to explore this distinction further (as there are other differences between nulls and overts besides the PCI/GCI distinction postulated here).

Furthermore, the disparity in interpretation preference between L1 and L2 participants in this experiment compared to smaller group differences in the previous studies on Italian and other European languages suggests that the two L2 groups face different tasks in acquiring an native-like interpretation pattern. For learners of Japanese, suppressing their English (same-subject) default interpretation for the overt pronoun is key, rather than shifting to an L2-like default interpretation for the pronoun, as in Italian and similar languages. This analysis is supported by the processing data, which shows L2 participants initially considering topic-shift at the pronoun, but then quickly falling back on an English-like interpretation at the verb. Conversely, once the English-like reaction is suppressed, L2 speakers shift completely to a Japanese-like interpretation pattern, without the “residual optionality” found in Italian studies, as
evidenced by the fact that L2 participants in this experiment either had interpretation patterns identical to L1 participants (three outliers) or completely different (everyone else).
REFERENCES


### LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACC</td>
<td>accusative</td>
</tr>
<tr>
<td>AP</td>
<td>against prediction</td>
</tr>
<tr>
<td>AQ</td>
<td>after question</td>
</tr>
<tr>
<td>COMP</td>
<td>complementizer</td>
</tr>
<tr>
<td>CONJ</td>
<td>conjunction</td>
</tr>
<tr>
<td>DAT</td>
<td>dative</td>
</tr>
<tr>
<td>DO1</td>
<td>direct object 1</td>
</tr>
<tr>
<td>DO2</td>
<td>direct object 2</td>
</tr>
<tr>
<td>ET</td>
<td>eye-tracking</td>
</tr>
<tr>
<td>FSMC</td>
<td>Form-Specific Multiple-Constraints</td>
</tr>
<tr>
<td>GCI</td>
<td>generalized conversational implicature</td>
</tr>
<tr>
<td>GEN</td>
<td>genitive</td>
</tr>
<tr>
<td>IH</td>
<td>Interface Hypothesis</td>
</tr>
<tr>
<td>JLPT</td>
<td>Japanese Language Proficiency Test</td>
</tr>
<tr>
<td>N</td>
<td>noun</td>
</tr>
<tr>
<td>NEG</td>
<td>negative</td>
</tr>
<tr>
<td>NOM</td>
<td>nominative</td>
</tr>
<tr>
<td>NOMI</td>
<td>nominalizer</td>
</tr>
<tr>
<td>NP</td>
<td>noun phrase</td>
</tr>
<tr>
<td>PAH</td>
<td>Position of Antecedent Hypothesis</td>
</tr>
<tr>
<td>PCI</td>
<td>particularized conversational implicature</td>
</tr>
<tr>
<td>PL</td>
<td>plural</td>
</tr>
<tr>
<td>PR</td>
<td>pronoun</td>
</tr>
<tr>
<td>PROG</td>
<td>progressive</td>
</tr>
<tr>
<td>PVT</td>
<td>picture verification task</td>
</tr>
<tr>
<td>Q</td>
<td>question</td>
</tr>
<tr>
<td>RT</td>
<td>reaction time</td>
</tr>
<tr>
<td>s.d.</td>
<td>standard deviation</td>
</tr>
<tr>
<td>SFP</td>
<td>sentence final particle</td>
</tr>
<tr>
<td>SS</td>
<td>same-subject</td>
</tr>
<tr>
<td>TOP</td>
<td>topic</td>
</tr>
<tr>
<td>TS</td>
<td>topic-shift</td>
</tr>
<tr>
<td>V1</td>
<td>verb 1</td>
</tr>
<tr>
<td>V2</td>
<td>verb 2</td>
</tr>
<tr>
<td>VOL</td>
<td>volitional</td>
</tr>
<tr>
<td>WP</td>
<td>with prediction</td>
</tr>
</tbody>
</table>