THE ACQUISITION AND ONLINE PROCESSING OF ANAPHORA
BY CHINESE-ENGLISH BILINGUALS: A COMPUTER ASSISTED STUDY

by

Rong Liu

________________________________________
Copyright © Rong Liu 2010

A Dissertation Submitted to the Faculty of the

GRADUATE INTERDISCIPLINARY DOCTORAL PROGRAM IN
SECOND LANGUAGE ACQUISITION AND TEACHING

In Partial Fulfillment of the Requirements
For the Degree of

DOCTOR OF PHILOSOPHY

In the Graduate College

THE UNIVERSITY OF ARIZONA

2010
As members of the Dissertation Committee, we certify that we have read the dissertation prepared by Rong Liu entitled *The Acquisition and Online Processing of Anaphora by Chinese-English Bilinguals: A Computer Assisted Study*

and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy

Janet Nicol

Robert Ariew

Kenneth Forster

Final approval and acceptance of this dissertation is contingent upon the candidate’s submission of the final copies of the dissertation to the Graduate College.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

Dissertation Director: Janet Nicol

Dissertation Director: Robert Ariew
STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at the University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgment of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the copyright holder.

SIGNED: Rong Liu
ACKNOWLEDGMENTS

I owe my deepest gratitude to all those people who have made this dissertation possible. I would like to begin by thanking my advisor, Dr. Janet Nicol, without whose valuable mentoring, tireless encouragement, and painstaking editing, this dissertation would not have been completed. Her insightful comments and constructive criticisms at different stages of my project helped me maintain a high research standard. She has introduced me to the field of psycholinguistics and inspired me to work on topics that I have never done before.

I have been very fortunate to have Dr. Robert Ariew as my dissertation co-chair. His expertise in language technology and pedagogy helped shape my project. I am deeply grateful to him for designing the template for the language training project. He has been always there to listen and give timely feedback. I am also thankful to him for the long discussions that helped me understand my research area better.

I am also indebted to Dr. Kenneth Forster, who is one of the best researchers and teachers that I have met in my life. He sets high standards for his students and provides guidance to meet those standards. He is always there to help and explain any question I have. I am grateful to him for his continuous encouragement and guidance.

I am thankful to my friend Dena McDuffie for carefully proofreading my manuscript and encouraging the use of correct grammar and consistent notation in my writings.

I would like to acknowledge my participants. They spent their valuable time with me. Without their participation and patience, my project would not have been finished.

Many friends have helped me during my graduate studies. Their generous support and care helped me stay focused on my study. I greatly appreciate their friendship.

I also wish to thank my family, a constant source of love, support and strength. My parents taught me how to be positive and persistent during difficult times. My wife helped me recruit participants and graciously assumed responsibilities such as housekeeping, cooking and other errands. My daughter always makes me smile.

Finally, I appreciate the financial support from the Language Learning Dissertation Grant, a dissertation grant from the Social and Behavioral Sciences Research Institute and a grant from the Asian American Faculty, Staff and Alumni Association at the University of Arizona.
DEDICATION

To my beloved wife Zoe and my daughter Hope.
TABLE OF CONTENTS

LIST OF FIGURES ....................................................................................................... 9
LIST OF TABLES ....................................................................................................... 10

ABSTRACT ................................................................................................................. 12

CHAPTER 1: INTRODUCTION ................................................................................ 13
  1.1 Introduction ........................................................................................................ 13
  1.2 Key Concepts ..................................................................................................... 16
    1.2.1 Anaphora ...................................................................................................... 16
    1.2.2 Implicit Knowledge versus Explicit Knowledge ......................................... 17
    1.2.3 Transfer ........................................................................................................ 17
    1.2.4 Input Processing ........................................................................................... 17
    1.2.5 Processing Instruction .................................................................................. 18
    1.2.6 CALL ........................................................................................................... 19
    1.2.7 Feedback in SLA .......................................................................................... 19
  1.3 Research Questions ............................................................................................ 19
  1.4 Significance of the Study.................................................................................... 21
  1.5 Conclusion .......................................................................................................... 22

CHAPTER 2: LITERATURE REVIEW ..................................................................... 23
  2.1 Introduction ........................................................................................................ 23
  2.2 First Language Online Pronominal Processing .................................................. 23
    2.2.1 Structural Constraints: The Binding Theory ................................................ 25
    2.2.2 Featural Constraints: Gender and Number Agreement ................................ 28
    2.2.3 Other Constraints/Cues ................................................................................ 30
  2.3 Second Language Anaphora Acquisition Studies ............................................... 31
    2.3.1 Offline Studies ............................................................................................. 31
    2.3.2 “Online” Studies .......................................................................................... 33
    2.3.3 Intervention Studies ..................................................................................... 35
  2.4 Linguistic Differences between English and Chinese Binding Principles ............. 38
  2.5 Processing Instruction (PI) ................................................................................ 40
    2.5.1 Features of Processing Instruction ............................................................... 41
    2.5.2 Differences between Traditional Instruction and Processing Instruction .... 43
    2.5.3 Principles of Processing Instruction ............................................................. 44
    2.5.4 Efficacy of Processing Instruction ............................................................... 46
    2.5.5 Rationale for Choosing PI as the Treatment ................................................ 48
  2.6 Explicit Feedback versus Implicit Feedback in SLA ............................................ 50
    2.6.1 Feedback in SLA .......................................................................................... 50
    2.6.2 Feedback in PI .............................................................................................. 54
  2.7 Conclusion .......................................................................................................... 57
# TABLE OF CONTENTS - Continued

CHAPTER 3: THE CURRENT STUDY ................................................................. 58
  3.1 Overview of the Current Study .......................................................... 58
  3.2 Research Questions and Hypotheses ............................................... 58
  3.3 Methodology .................................................................................... 60
    3.3.1 Online Methods versus Offline Methods .................................. 60
    3.3.1.1 Traditional Method: Grammaticality Judgment ..................... 61
    3.3.1.2 Online Method: Self-Paced Moving Window Reading Tasks .... 63
    3.3.2 Computer-Delivered Processing Instruction ............................... 65
  3.4 Conclusions .................................................................................... 66

CHAPTER 4: EXPERIMENT 1 ......................................................................... 68
  4.1 Introduction ..................................................................................... 68
  4.2 Research Questions and Hypotheses ............................................... 68
  4.3 Methodology .................................................................................... 70
  4.4 Instrumentation ............................................................................... 71
    4.4.1 Self-Paced Reading Tasks ....................................................... 71
    4.4.2 Offline Written Test ................................................................. 73
    4.4.3 Language Profile ..................................................................... 75
  4.5 Participants ..................................................................................... 75
  4.6 Procedures ..................................................................................... 77
  4.7 Data Analysis and Results ............................................................... 78
    4.7.1 Reading Time Patterns at Positions 1 and 2 ............................... 82
    4.7.2 Reading Time Patterns at Positions 3 and 4 ............................... 83
  4.8 Discussion and Conclusion ............................................................. 85

CHAPTER 5: EXPERIMENT 2 ......................................................................... 88
  5.1 Introduction ..................................................................................... 88
  5.2 Research Questions and Hypotheses ............................................... 88
  5.3 Methodology .................................................................................... 90
  5.4 Instrumentation ............................................................................... 90
    5.4.1 Self-Paced Reading Tasks ....................................................... 90
    5.4.2 Offline Written Test ................................................................. 93
    5.4.3 Language Profile ..................................................................... 95
  5.5 Participants ..................................................................................... 95
  5.6 Procedures ..................................................................................... 96
  5.7 Data Analysis and Results ............................................................... 97
    5.7.1 Reading Time Patterns at Positions 1 and 2 ............................... 101
    5.7.2 Reading Time Patterns at Positions 3 and 4 ............................... 101
  5.8 Discussion and Conclusion ............................................................. 104
CHAPTER 6: EXPERIMENT 3 ................................................................................ 108
  6.1 Introduction ...................................................................................................... 108
  6.2 Research Questions and Hypotheses ................................................................. 108
  6.3 Methodology .................................................................................................... 110
  6.4 Instrumentation ............................................................................................... 111
    6.4.1 Self-Paced Reading Tasks: Pretest and Posttest ........................................... 111
    6.4.2 Offline Written Test: Pretest and Posttest ............................................... 112
    6.4.3 Language Profile ...................................................................................... 115
    6.4.4 Computer-Delivered Processing Instruction: Treatment ........................... 115
    6.4.5 Exit Survey of the Treatment .................................................................. 125
  6.5 Participants ..................................................................................................... 126
  6.6 Procedures ....................................................................................................... 127
  6.7 Data Analysis and Results .............................................................................. 129
  6.8 Discussion and Conclusion .......................................................................... 140

CHAPTER 7: DISCUSSION AND CONCLUSIONS .............................................. 144
  7.1 Introduction .................................................................................................... 144
  7.2 Synopsis of Main Findings and Discussion ..................................................... 144
    7.2.1 Experiment 1 ............................................................................................ 145
    7.2.2 Experiment 2 ............................................................................................ 153
    7.2.3 Experiment 3 ............................................................................................ 157
  7.3 Implications for SLA and Language Pedagogy ................................................ 160
  7.4 Limitations of the Present Study ...................................................................... 165
  7.5 Suggestions for Future Research .................................................................... 166
  7.6 Conclusion ..................................................................................................... 168

APPENDIX A: ONLINE TEST STIMULI IN EXPERIMENT 1 ......................... 170
APPENDIX B: OFFLINE TEST IN EXPERIMENT 1 .......................................... 176
APPENDIX C: LANGUAGE PROFILE ................................................................. 178
APPENDIX D: RESULTS OF REACTION TIMES IN EXPERIMENT 1 ............... 180
APPENDIX E: ONLINE TEST STIMULI IN EXPERIMENT 2 ............................ 182
APPENDIX F: OFFLINE TEST IN EXPERIMENTS 2 AND 3 ............................. 186
APPENDIX G: RESULTS OF REACTION TIMES IN EXPERIMENT 2 ............ 189
APPENDIX H: OFFLINE TEST IN EXPERIMENT 3 .......................................... 191
APPENDIX I: COMPUTER-DELIVERED PROCESSING INSTRUCTION IN   
  EXPERIMENT 3 ................................................................................................. 198
APPENDIX J: EXIT SURVEY IN EXPERIMENT 3 ............................................ 207
APPENDIX K: RESULTS OF REACTION TIMES IN EXPERIMENT 3 .............. 209

REFERENCES ........................................................................................................ 211
LIST OF FIGURES

Figure 1: L2 acquisition processes according to PI ..................................................... 43

Figure 2: Traditional practice ....................................................................................... 43

Figure 3: Screen shot: Log in ..................................................................................... 120

Figure 4: Screen shot: Instructions for activity one ................................................... 121

Figure 5: Screen shot: After audio stimulus for activity one ..................................... 122

Figure 6: Screen shot: Feedback ................................................................................ 122

Figure 7: Screen shot: End of the section ................................................................. 123

Figure 8: Screen shot: Explicit feedback ................................................................. 124

Figure 9: Screen shot: Beginning of the survey ......................................................... 125

Figure 10: Screen shot: Sample survey question ....................................................... 126

Figure 11: Scores of the interpretation task in the pretest and posttest for each
individual in the implicit feedback group (The mean for pretest M=0.59; and the mean
for posttest M=0.93) .................................................................................................. 130

Figure 12: Scores of the interpretation task in the pretest and posttest for each
individual in the explicit feedback group (The mean for pretest M=0.56; and the mean
for posttest M=0.90) .................................................................................................. 130

Figure 13: Scores of the production task in the pretest and posttest for each individual
in the implicit feedback group (The mean for pretest M=0.98; and the mean for
posttest M=0.95) ........................................................................................................ 131

Figure 14: Scores of the production task in the pretest and posttest for each individual
in the explicit feedback group (The mean for pretest M=0.93; and the mean for
posttest M=0.93) ........................................................................................................ 131

Figure 15: Mechanism of coreference online processing ...................................... 155
**LIST OF TABLES**

Table 1: Comprehension error rates (percentage of errors per sentence type) ...............34

Table 2: Average response time in seconds per sentence type........................................35

Table 3: Nonnative speakers’ language profile in Experiment 1.................................77

Table 4: Native speakers mean RTs (ms) at four test positions of an example sentence for grammatical and ungrammatical sentences involving number cues and gender cues..............................................................................................................80

Table 5: Native speakers mean RTs at four test positions in example sentences for plausible and implausible sentences involving verb cues..............................................80

Table 6: Nonnative speakers mean RTs (ms) at four test positions for grammatical and ungrammatical sentences involving number cues and gender cues..............................81

Table 7: Nonnative speakers mean RTs at four test positions for plausible and implausible sentences involving verb cues.................................................................81

Table 8: Nonnative speakers’ language profile in Experiment 2.................................96

Table 9: Native speakers’ mean RTs (ms) at four test positions for neutral sentences and pragmatically biased sentences involving the structural cue for conditions “both match” and “no match” .................................................................................................99

Table 10: Native speakers’ mean RTs (ms) at four test positions for neutral sentences and pragmatically biased sentences involving the structural cue for conditions “only NP1 match” and “only NP2 match” ..............................................................................................................99

Table 11: Nonnative speakers’ mean RTs (ms) at four test positions for neutral sentences and pragmatically biased sentences involving the structural cue for conditions “both match” and “no match” ..................................................................................100

Table 12: Nonnative speakers’ mean RTs (ms) at four test positions for neutral sentences and pragmatically biased sentences involving the structural cue for conditions “only NP1 match” and “only NP2 match” ..............................................................100

Table 13: Participants’ background in Experiment 3..................................................127
LIST OF TABLES - Continued

Table 14: Participants’ mean RTs (ms) at four test positions for neutral sentences and pragmatically biased sentences involving the structural cue for conditions “both match” and “no match” in the pretest ................................................................. 134

Table 15: Participants’ mean RTs (ms) at four test positions for neutral sentences and pragmatically biased sentences involving the structural cue for conditions “only NP1 match” and “only NP2 match” in the pretest ................................................................. 134

Table 16: Participants’ mean RTs (ms) at four test positions for neutral sentences and pragmatically biased sentences involving the structural cue for conditions “both match” and “no match” in the posttest ................................................................. 135

Table 17: Participants’ mean RTs (ms) at four test positions for neutral sentences and pragmatically biased sentences involving the structural cue for conditions “only NP1 match” and “only NP2 match” in the posttest ................................................................. 135
ABSTRACT

This dissertation investigated the acquisition and processing of anaphora by learners of English, using both “offline” methods such as paper-and-pencil written tests and “online” methods such as self-paced reading-for-comprehension. Three experiments were conducted. The first two experiments tested advanced Chinese ESL learners’ knowledge and processing of anaphora. The focus of the first experiment was on whether participants automatically use gender and number cues, and verb information to interpret pronouns and reflexives during online processing. The second experiment manipulated gender and pragmatic cues to test whether participants have acquired knowledge of structural constraints on reflexive interpretation (i.e., the binding principles). The third experiment, using a pretest-treatment-posttest design, investigated the efficacy of computer-delivered Processing Instruction (PI) on the acquisition of structural constraints and the use of those constraints during reading. During the pretest, subjects completed a self-paced reading task and a written test. For the treatment, participants learned the grammatical constraints on reflexives through interaction with a computer program. Posttest assessment included one interpretation test, one sentence completion task, and one self-paced reading task. The role of feedback in Computer Assisted Language Learning was also examined. Results showed the following: (1) Advanced L2 learners were more sensitive to certain types of agreement information (gender) than others (number). (2) PI improved L2 learners’ knowledge about constraints on reflexives as measured by offline tests. (3) PI led to improvement in learners’ processing strategies as measured by online tasks. (4) No significant difference was found between the implicit feedback group and the explicit feedback group in the third experiment. Overall, this research highlights the importance of multiple types of assessment that tap the acquisition of grammatical knowledge as well as the proficiency with which learners use that knowledge during reading comprehension tasks.
CHAPTER 1: INTRODUCTION

1.1 Introduction

Online grammatical processing has been one of the major areas of first language (L1) acquisition. Psycholinguists have employed a variety of advanced techniques such as self-paced reading, eye-tracking and brain-imaging to study strategies that L1 speakers use when understanding a sentence. Researchers in the second language acquisition field have only recently started to use those methods to study second language (L2) grammatical processing (Felser, 2005). Juffs (2001) once acknowledged that it is a little embarrassing that reaction time studies have been used by L2 researchers “only in the past dozen years or so” (Juffs, 2001, p.208). Many researchers (N. Ellis, 1993; Hulstijin, 2001; Segalowitz, 2003; Segalowitz, Segalowitz, and Wood, 1998) agree that the ultimate goal of L2 learning and teaching is to cultivate the ability to use the language spontaneously and automatically. Unfortunately, as every English as a Second Language (ESL) teacher or learner can tell, L2 processing is generally slower and less automatic than mature L1 processing. And hence it is less native-like.

One possible reason that L2 processing is nonnative-like is the inappropriate processing strategies used during online sentence comprehension (Marinis, 2003). According to Marinis (2003), failure to acquire the target language’s processing strategies can result in nonnative-like attainment in L2. Others such as Clahsen and
Felser (2006) proposed the Shallow Structure Hypothesis and claimed that L2 sentence processing is fundamentally different from L1 processing. They claimed that L2 processing is syntactically shallower and learners rely more on lexical and pragmatic information than syntactic information when understanding sentences. Another possible reason is that there are different types of knowledge: implicit and explicit. The type of knowledge that L1 speakers use for spontaneous and fluent speech and reading is implicit knowledge. While explicit knowledge is often learned through intentional learning or instruction. Of course, there is still considerable controversy over the question of how to distinguish implicit and explicit knowledge precisely. However, it is believed that online experimental psycholinguistic methods enjoy the advantage that they can tap into the implicit knowledge that a speaker possesses. This is the very reason that online psycholinguistic methods will be employed in this study. This topic will be picked up again in Chapter 2.

One of the major questions that are of interest to L2 researchers is whether L2 speakers can process sentences as L1 speakers and what factors may contribute to their ability to process sentences as L1 speakers do. This study intends to investigate whether L2 readers process pronouns and reflexives in the same way as L1 speakers and what constraints or cues are used during their online processing. Anaphora is a perfect target structure because it is well studied by L1 processing researchers. They found that to understand anaphora, readers need to use multiple cues in the sentence which will be discussed in detail in the literature review. And to my knowledge, this
study is among the first studies that use psycholinguistic methods to study L2 anaphora processing. In addition, the study intends to examine the efficacy of training on anaphora processing strategies by computer-delivered processing instruction (PI). Many studies (e.g., VanPatten, 2004) have found that processing instruction, an input-based pedagogical technique, is superior to traditional instruction. PI is one of the few teaching methods that take learners’ psycholinguistic strategies into consideration for the purpose of instruction. However, it is still inconclusive whether processing instruction can lead to gains in the development of implicit knowledge because previous studies used only offline methods to measure the outcome, which may just capture the gains in explicit knowledge. This study will use both offline methods and online methods to measure the effectiveness of the treatment. Therefore both implicit knowledge and explicit knowledge will be measured.

Another major area of the study is the role of feedback in second language acquisition (SLA). What types of feedback benefit learners most? Should teachers provide corrective feedback whenever possible? Should materials in Computer Assisted Language Learning (CALL) include feedback? There are conflicting results in the literature about those questions. In general, studies in SLA pointed to an advantage of explicit feedback over implicit feedback. However, recent studies in PI found that the type of feedback did not matter as long as the instructional activities are structured as posited in the principles of PI. The study will test the role of different
types of feedback in PI.

There are three experiments in the study. The first two will identify L2 speakers’ processing strategies of comprehending anaphora as sentences unfold over time. The major purpose is to document the mechanism that L2 learners use to process pronouns and reflexives during their real-time sentence comprehension. And the third experiment will use computer-delivered processing instruction to alter participants’ inappropriate processing strategies when reading anaphora. It aims to test whether PI is effective in grammar teaching and to investigate the role of feedback.

1.2 Key Concepts

The study works at the intersection of pedagogy and psycholinguistics. It involves several key concepts that are important for the conceptualization of this research. They are: anaphora, implicit knowledge and explicit knowledge, transfer, input processing, processing instruction, CALL, feedback in SLA. The following will define and discuss those key concepts briefly in the study.

1.2.1 Anaphora

Anaphora in this study is used as a general term and it refers to pronouns such as “he”, “she”, and “they” and reflexives such as “himself”, “herself”, and “themselves”. It does not include reciprocals such as “each other”.

1.2.2 Implicit Knowledge versus Explicit Knowledge

Implicit knowledge refers to the knowledge that speakers use to do things automatically that they may or may not be able to verbalize explicitly. It is known as “knowing how”. Explicit knowledge is “knowing that” and refers to knowledge that speakers are aware of and may be able to express explicitly.

1.2.3 Transfer

Transfer refers to the native language influence on the acquisition of L2. It can be positive or negative. The general consensus is that L1 does shape L2 acquisition but it is not the only source.

1.2.4 Input Processing

Input processing (IP) was first proposed by VanPatten (1996). It concerned the way learners perceive and process input. According to VanPatten, learners filter input (i.e., only process certain forms of input). Those linguistic data that learners attend to and hold in their working memories during online comprehension are called intake (VanPatten, 2004). Psycholinguistic strategies and mechanisms that learners use to derive intake from input play an important role in IP. VanPatten (1996, 2003, 2004) presented a set of principles of IP. One principle, the Primacy of Meaning Principle, claimed that during comprehension, learners try every means to understand the meaning from the input. Thus, content words are searched out first. When both content words and grammatical forms encode the same referential meaning, learners
prefer to rely on lexical items. Another notion, communication value, is important in IP. It refers to “the meaning that a form contributes to overall sentence meaning and is based on two features: [+/– inherent semantic value] and [+/– redundancy]” (VanPatten, 2004, p.139). The communicative value is greater if the features are positive semantic value and not redundant. If the semantic value is negative, the communication value is zero. For example, in “Jack walked the dog yesterday”, the past tense morpheme –ed was redundant because the semantic value is realized elsewhere by “yesterday”.

IP attempts to describe what learners do when they encounter input online and it includes learners’ psycholinguistic strategies in the model. It is still under development and the sketch here is brief. Nevertheless, it provides enough background for the discussion of processing instruction, a pedagogical technique based on IP, in the current study.

1.2.5 Processing Instruction

Processing instruction (PI) is a pedagogical application of input processing (IP), which refers to how learners get intake from input when processing a language. Intake refers to those “linguistic data actually processed from the input and held in working memory for further processing” (VanPatten, 2005, p.268). Intake does not equal acquisition. Further discussion about PI will occur in Chapter 2.
1.2.6 CALL

CALL is the abbreviation for computer-assisted language learning. In this study, the computer takes the role of a “tutor” in the instruction and therefore it is classified as tutorial CALL. Although tutorial CALL has been stigmatized because it is considered to be associated with behaviorism, recent studies (e.g., Hubbard and Bradin-Siskin, 2004) showed that it is not necessarily true because learners could be autonomous. In addition, the need for tutorial CALL materials has increased (Howell, Williams, and Lindsay, 2003) due to the high demands of distance education and the demands of supplemental language learning materials by teachers and learners.

1.2.7 Feedback in SLA

Feedback in the study refers to the linguistic data or information given to a learner following an erroneous language behavior. Feedback can be implicit such as recasts and explicit such as providing explicit grammatical rules.

1.3 Research Questions

One purpose of the study is to find out how Chinese learners of English understand pronouns and reflexives in sentences. For example, in “The mother said that the daughter was preparing herself for the party”, English speakers know that “herself” must mean or “co-refer with” “daughter”, but in Chinese sentences with the same structure, “herself” can also mean “mother”. Therefore, it is difficult for Chinese
learners of English to correctly interpret such sentences. Another purpose of this study is to use computer-assisted language learning techniques to investigate whether “Processing Instruction” (VanPatten and Cadierno, 1993), a popular pedagogical teaching tool, is effective in developing learners’ comprehension abilities. Specifically, the research questions are:

1. Can L2 learners understand pronouns (e.g., “he”, “she”, and “they”) and reflexives (e.g., “himself”, “herself”, and “themselves”) in the same way as native language (L1) speakers? For example, will they be able to take the structure of the sentence into account in interpreting these elements? In a sentence such as “Bill said that Clark hates himself” “himself” cannot refer to “Bill” because “Bill” is in the wrong position in the structure. Likewise, in sentences such as “Bill said that Clark hates him”, “him” cannot mean “Clark” because “Clark” is part of the same clause as the reflexive. These illustrate the “structural constraints on co-reference”. Can learners use gender and number information to interpret pronouns and reflexives? For example, “Mary said that the boys hate herself” is ungrammatical because “the boys” and “herself” do not match in gender and number. Can learners use verb cues? For example, in “Jane admires Jill because she is reliable”, readers link “she” with the object “Jill” instead of the subject “Jane” because the verb “admire” usually takes the object as the causer of the action.

2. Can we help learners develop the correct processing routines through computer-delivered Processing Instruction?
3. What kind of feedback benefits learners best during the processing instruction? Should we just tell them “right”/“wrong” or should we provide grammatical rules as well?

1.4 Significance of the Study

There is a scarcity of L2 studies using psycholinguistic methods to study grammatical processing. This study will fill in that gap. In addition, the study is a unique combination of psycholinguistics and pedagogy, i.e., using psycholinguistic methods to answer pedagogical questions. Using psycholinguistic methods, the study can help us understand the mechanisms that underlie L2 grammatical processing and how they differ from L1 processing, and tell us to what extent learners can acquire the processing strategies as employed by native speakers. It will explore the usefulness of psycholinguistic methods in SLA. It will also test whether processing instruction, a focus-on-form pedagogical technique, can alter L2 learners’ anaphora processing strategies. Previous studies have only attested to its effectiveness by traditional offline measures such as written tests. To better control intervening variables, this study will employ computers, instead of instructors, to deliver the instruction. Few PI studies have done so in the past. The role of feedback in CALL will be investigated and it will inform us what type of feedback is most effective. The results will have practical implications for language instruction and the design of CALL materials.
1.5 Conclusion

This chapter introduced the key issues in second language acquisition (SLA) that inspired the current research. Answers to those questions can have deep theoretical and practical impact. Theoretically, it will shed light on the mechanism under which a learner processes and acquires a language. Practically, it will inform us what pedagogical technique is effective and what kind of feedback teachers should provide to learners.

Chapter 2 critically reviews the relevant literature informing the major issues introduced in Chapter 1, including first language online pronominal processing, second language anaphora acquisition, cross-linguistic differences in binding principles, processing instruction and the role of feedback in SLA. Chapter 3 presents an overview of the current study, including specific research questions and hypotheses and the research methods that were used to gather data. The next three chapters, i.e., Chapter 4, 5, and 6, give a detailed account of one experiment that was designed to answer the research questions, including methodology, instrumentation, data collection procedures, data analysis, results, and a brief discussion. Finally, Chapter 7 takes a more detailed look at the findings, interprets and discusses the theoretical and practical implications for L2 pedagogy as well as the limitations of the study, and offers suggestions for further study.
CHAPTER 2: LITERATURE REVIEW

2.1 Introduction

This chapter reviews the relevant literature, including studies on first language (L1) online pronominal processing, second language (L2) anaphora acquisition studies, English and Chinese binding principles, processing instruction and the role of feedback in SLA and PI. It lays out the theoretical foundation for the current study.

2.2 First Language Online Pronominal Processing

Reference resolution is vital to language comprehension. In order to interpret sentences correctly, anaphoric expressions must be “coindexed with”, or linked to, their referents. Although there are many types of anaphoric expressions that have been the foci of psycholinguistic study of anaphora, my study will be limited to a subset of these: pronouns such as “he”, “she”, and “they” and reflexives such as “himself”, “herself” and “themselves”.

According to Nicol and Swinney (1989, 2002), during online sentence comprehension, initial anaphoric processing may trigger the search for an antecedent. A candidate set of antecedents will be considered and all but one antecedent will be eliminated. Therefore, coreference process is three-staged: Stage 1: anaphora triggers the search for an antecedent; Stage 2: a set of candidates which includes aforementioned noun phrases are considered; Stage 3: non-antecedent candidates are
eliminated and one remains. For example, when a reader read the sentence “Bill says that John hates himself on TV”, the appearance of “himself” triggers the search process of an antecedent. Then, the two previously mentioned NPs, “Bill” and “John” enter the candidate set. Finally, “Bill” is eliminated and “John” is chosen as the antecedent. This multiple-staged coreference processing captures the general mechanism of anaphoric processing during online sentence comprehension.

One focus of research on the comprehension of anaphora in L1 has to do with what constrains the process of anaphora resolution in L1 during online sentence processing. Using various methods such as self-paced reading, eye-tracking during reading, probe verification and cross-modal semantic priming, previous research has uncovered a large number of factors affecting the processing of pronouns and reflexives. The major ones are: 1. Structural constraints, i.e., the role of syntactic structure on coreference processing. 2. Featural constraints, i.e., the role of pronominal features such as gender, number and animacy (or humanness) on the interpretation of anaphora. 3. Other constraints: a. Lexical and pragmatic constraints, including “implicit causality” and world knowledge on the processing of pronouns and reflexives. b. Heuristic strategies such as the first-mention preference, and subject preference, and parallel function strategy. c. Repeated-name penalty, i.e., repeated names caused increased reading times compared to pronouns (Gordon, Grosz and Gilliom, 1993). The focus of this dissertation will be on constraints 1 and 2 and implicit causality associated with verbs in two-clause sentences. Other constraints will
not be discussed.

2.2.1 Structural Constraints: The Binding Theory

It is generally agreed that sentence structure plays a critical role in coreference assignment. Chomsky (1981) outlined three well-known principles in his binding theory:

   Principle A: An anaphor (reflexives such as “himself” and reciprocals such as “each other”) must be bound in its local domain (for the present purpose, the local domain may be interpreted as the minimal clause containing the anaphor and a subject)

   Principle B: A pronoun is free in a local domain.

   Principle C: An R-expression (such as proper names and other full referring expressions) is free.

   In the following examples, underlined words of sentences in 1a, 1b, and 1c corefer while those in (1’) cannot corefer because each violates the principle a, b, and c respectively.

   (1) a. Mary hates herself. (Principle A)

   b. Bill said that Clark hates him. (Principle B)

   c. Jane said that Mary hates her. (Principle C)

   (1’) a. Bill said that Mary hates himself.

   b. Bill said that Clark hates him.

   c. She said that Mary hates dogs.
With respect to online comprehension of sentences of this type, it is unclear when this constraint comes into play. While most researchers agree with the existence of structural constraints on coreference assignment, they disagree with the exact nature of the constraints, especially concerning the time course of those constraints on the initial online coreference processing and its interaction with other constraints. However, in this study it is assumed that the binding principles are applied at the earliest stage and only structural accessible candidates are considered as possible antecedents, i.e., the structural constraints function as an initial filter for the initial antecedent candidates. This view was first advanced by Nicol and Swinney (1989). They presented subjects with sentences such as examples 2a and 2b to probe whether and when the noun phrases preceding a pronoun or reflexive are considered as possible candidates.

(2)  

   a. The boxer told the skier that the doctor for the team would blame himself for the recent injury.

   b. The boxer told the skier that the doctor for the team would blame him for the recent injury.

Subjects were asked to make a lexical decision to semantic associates of the three preceding NPs (boxer, doctor, skier), which were visually presented on a screen when subjects heard the pronoun or the reflexive in the above sentences from audio stimuli. Priming (compared to control words matched in frequency and length) was found only for the structurally permissible antecedent such as “the doctor” in example 2a,
but not for the structurally inaccessible antecedents “the boxer” and “the skier”. The same pattern was obtained for 2b, i.e., no priming for “doctor”, but significant priming for both “boxer” and “skier”. The results led them to conclude that the initial set of candidates is restricted by grammatical constraints and they maintained that “The initial set of candidate antecedents contains all and only those referents that bear the appropriate syntactic relation to the referentially dependent NP” (Nicol and Swinney, 1989, p.17).

Additional evidence comes from Clifton, Kennison and Albrecht (1997). They used self-paced reading tasks to test the following sentences to investigate if binding theory acts as an initial filter.

(3) a. The supervisors paid him yesterday to finish typing the manuscript.

b. The supervisor paid him yesterday to finish typing the manuscript.

c. The supervisors paid his assistant yesterday to finish typing the manuscript.

d. The supervisor paid his assistant yesterday to finish typing the manuscript.

They predicted that if the binding principles initially filter out antecedents, there should be a match/mismatch effect (i.e., a reading slowdown occurs when an anaphoric expression does not match the features such as gender, number of an NP) in examples 3c and 3d but not in 3a and 3b. The reading times of the region following the pronoun were longer in 3c compared to the other three sentences. This was taken as evidence that binding principles serve as an initial filter and hence support the initial-filter model.
2.2.2 Featural Constraints: Gender and Number Agreement

There is no doubt that gender, number and animacy features of an antecedent must agree with those of the anaphoric expressions such as pronouns and reflexives for the final coreference resolution. The question is when featural information is available to the processor. Conflicting results are reported. Some research finds immediate effects: potential antecedents are considered only if they match the anaphoric element in terms of features such as gender, number and animacy (e.g., Nicol, 1988); features are not automatically used (e.g., Greene, McKoon and Ratcliff, 1992). In this study it is assumed that featural information is used immediately and automatically by L1 speakers. Evidence for automatic gender use was reported by Nicol (1988). The following sentences were tested:

(4) a. The ballerina told the skier that the doctor for the team would blame him for the recent injury.

b. The ballerina told the skier that the doctor for the team would blame her for the recent injury.

Using semantic associates as probes, significant priming was found for only those antecedents that did not mismatch in gender with pronouns. So only for the “him” condition 4a, there was priming for “skier” while for the “her” condition 4b, there was priming for both “ballerina” and “skier”. She also showed that number was automatically used by testing the sentences such as 5.

(5) The boxers told the skier that the doctor for the team would blame them/him for the injury.
Again priming was found only for those NPs that agreed in number with the pronoun.
So for the “them” condition, priming was found only for “boxers”, and for the “him”
condition, only for “skier”.

The idea of automatic gender use in pronoun resolution has also received support
from eye-tracking studies by Arnold, Eisenband, Schmidt and Trueswell (2000).
Subjects listened to passages for general comprehension while viewing cartoon
pictures with two characters of either same or different gender. Subjects’ eye
movements were recorded.

(6). Donald is bringing some mail to Minnie./ He’s sauntering down the hill,/ while
a violent storm is beginning./ She’s carrying an umbrella, /and it looks like they’re
both going to need it.

Subjects rapidly used the gender cue to interpret “she” in sentence (6) as “Minnie”
even when it referred to an antecedent that is not highly accessible (compared to the
first-mentioned character Donald).

Other evidence supporting an automatic gender pronominal process comes from
Rigalleau and Caplan (2000). Using a cross-modal naming paradigm (i.e., subjects
listened to the first part of a sentence and named a probe word that appeared on a
screen, after which they listened to the end of the sentence. The audio was stopped at
the probe word and continued after subjects named it.), they provided evidence that a
disagreement between a pronoun and the noun in the discourse focus disrupted
subjects’ processing. They tested the following sentences in their experiment 3 and
found the naming latency of pronouns was longer if there is disagreement between the
antecedent and the pronoun in gender. They concluded that there is immediate automatic gender coreference.

(7). Nancy paid without being asked; he had no sense of honor.

2.2.3 Other Constraints/Cues

Besides structural and featural constraints, other factors have also been identified to influence coreference processing. When there are multiple possible antecedent candidates, other factors come into play to identify a unique one. Among them, one well-studied factor is verb-based “implicit causality” (Garnham and Oakhill, 1985; McDonald and MacWhinney 1995). In causal constructions, verbs are biased toward NP1 or NP2 (Stewart, Pickering, and Sanford, 2000) as their antecedents. In “Jane admires Jill because she is reliable”, readers link “she” with “Jill” (NP2) instead of “Jane” (NP1) because the verb “admire” usually takes the object as the causer of the action. This type of verb is called Experiencer-Stimulus. “Bore” is another type of verb, i.e., Stimulus-Experiencer. For example, “Jane bores Jill because she is repetitious”. Here “she” refers to “Jill” (NP1). McDonald and MacWhinney (1995) found that when there is a conflict between the explicit and implicit causes, people take longer to process anaphors. For example, people take more time to read a sentence such as “Jane bores Jack because he is repetitious” than “Jane bores Jack because she is repetitious”. In this study the term “implicit causality” is used as a general term to refer to the phenomena that certain verbs such as “bore” prefer to take
a subject or an object such as “envy” as the causer of the action especially when two clauses are linked by “because” as the above examples showed. A violation of the preference of a verb does not necessarily render a sentence ungrammatical. For example, “Mike praised Jill because he was very proud”.

2.3 Second Language Anaphora Acquisition Studies

Most previous studies on anaphora acquisition are from the perspective of universal grammar (Hirakawa, 1990; Lee and Schachter, 1997, Wakabayashi, 1996; White, 1998). The major question examined was whether L2 learners could “reset the binding parameters”, i.e., when there is difference in the value of a parameter between L1 and L2 such as whether long-distance antecedent is allowed, learners need to change the value of the parameter for L2. Most studies show that L2 learners can acquire the binding principles but it is hard to reach ultimate attainment, i.e., acquire all properties of pronouns, for example, some learners whose L1 allows long-distance binding still allows long-distance antecedents in English (See Thomas, 1995 for a thorough review). The following will review some important offline and online studies.

2.3.1 Offline Studies

Offline L2 studies show that L2 learners may over-rely on semantic and pragmatic information in making grammaticality judgments and truth value
judgments. Thomas (1989) investigated the extent that pragmatics influences L2 learners’ reflexive assignment. In sentences such as “Mary angrily told me that Susan has spilled a lot of paint on herself”, the pragmatic meaning favors “Mary” as the antecedent because she was angry. In Chinese, both “Mary” and “Susan” can be “herself”. But the pragmatic meaning will favor “Mary” as the antecedent of “herself”. Of course, in English, pragmatics does not play a role in the interpretation here and only “Susan” can be the antecedent. Results showed that among Chinese learners many resisted a long-distance antecedent despite the pragmatic bias in those sentences. However, Demirci (2000) provided evidence that Turkish L2 learners of English appear to transfer pragmatics-based strategies from L1 to L2. Regardless of proficiency, they were misled by pragmatic bias. For example, they interpreted “himself” as “the little boy” in sentences such as “The little boy was angry that his father always hit himself”. This is also true when the bias is toward the local NP.

Besides using grammaticality judgment, truth value judgment was also used in the literature. In this task, subjects read a passage and then are asked to answer a true/false question according to the passage. Because how subjects interpret the question influences their answer, it is a good way to test learners’ grammar knowledge without explicitly asking them about it. Using a story-based truth value judgment task, Akiyama (2002) found evidence that Japanese learners of English had difficulty acquiring the locality principle especially for embedded infinitival clauses such as “John wanted Tom to know himself better”. They did better in sentences with
embedded “that clauses” such as “John thought that Tom was blaming himself”.

2.3.2 “Online” Studies

There is little research on the processing of L2 anaphora as a sentence unfolds. However, Cook (1990), using a computer-controlled timed comprehension task, studied advanced English learners’ interpretations of reflexives and pronominals such as “him”. In the experiment, participants were presented with a whole sentence such as “Peter said John helped himself.” and two forced choices, i.e., pressing A if “he/himself” refers to “Peter”; pressing B if it refers to “John”. The computer program recorded participants’ response and also the reaction time of each response. Subjects were 47 advanced English learners whose L1s were Japanese, Romance and Norwegian. Comprehension error data showed that overall, reflexives were slightly more difficult than pronominals with an error rate of 12.6% for reflexives and 10.1% for pronominals. And with few exceptions, there was a rather consistent order of difficulty for the five sentence types, i.e., sentence type H (See Table 1) was the most difficult; the simple sentence type with no embeddings (Type B) the least difficult. Reaction time data in Table 2 showed that there was a common order of response time for sentence types, i.e., see Table 1, questions BDF and H. This means that the easiest is the simple “John shot himself”, then the tensed “Peter said that John voted for himself”, then the infinitival “Peter asked John to include himself”, and the most difficult, the noun phrase “John reported Peter’s criticisms of himself”. Because in the
experiment sentence length is not constant, and response time includes both the reading time of the sentences and the judgment time, the reaction time data can only be a rough measure of the difficulty level of those structures. This study does not examine what learners do moment-by-moment when interpreting anaphora, i.e., informing us little about what cues or information are used to assign antecedents. The focus of the study again is whether L2 learners have access to UG. However, it informs us that different sentence structures pose different problems for learners. The study established that the resolution of anaphora depends on the complexity of the sentence containing the anaphoric expressions.

Table 1: Comprehension error rates (percentage of errors per sentence type)

<table>
<thead>
<tr>
<th></th>
<th>Native (n=14)</th>
<th>Nonnative (n=47)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Peter shot him.</td>
<td>10.8</td>
<td>7.4</td>
</tr>
<tr>
<td>B. John shot himself.</td>
<td>1.8</td>
<td>1.7</td>
</tr>
<tr>
<td>C. Peter said that John voted for him.</td>
<td>7.1</td>
<td>6.1</td>
</tr>
<tr>
<td>D. Peter said that John voted for himself.</td>
<td>8.9</td>
<td>11.1</td>
</tr>
<tr>
<td>E. John asked Peter to pay for him.</td>
<td>12.5</td>
<td>15.5</td>
</tr>
<tr>
<td>F. Peter asked John to include himself.</td>
<td>7.1</td>
<td>30.0</td>
</tr>
<tr>
<td>G. Peter discovered John’s report on him.</td>
<td>14.3</td>
<td>19.3</td>
</tr>
<tr>
<td>H. John reported Peter’s criticisms of himself.</td>
<td>16.0</td>
<td>37.7</td>
</tr>
<tr>
<td>I. Peter saw John save him.</td>
<td>10.7</td>
<td>17.8</td>
</tr>
<tr>
<td>J. Peter saw John save himself.</td>
<td>8.9</td>
<td>10.8</td>
</tr>
</tbody>
</table>

(Adapted from Cook, 1990)
Table 2: Average response time in seconds per sentence type

<table>
<thead>
<tr>
<th></th>
<th>Native (n=14)</th>
<th>Nonnative (n=47)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Peter shot him.</td>
<td>3.3</td>
<td>4.5</td>
</tr>
<tr>
<td>B. John shot himself.</td>
<td>3.0</td>
<td>3.8</td>
</tr>
<tr>
<td>C. Peter said that John voted for him.</td>
<td>4.6</td>
<td>6.8</td>
</tr>
<tr>
<td>D. Peter said that John voted for himself.</td>
<td>4.2</td>
<td>6.8</td>
</tr>
<tr>
<td>E. John asked Peter to pay for him.</td>
<td>4.7</td>
<td>7.1</td>
</tr>
<tr>
<td>F. Peter asked John to include himself.</td>
<td>4.5</td>
<td>7.1</td>
</tr>
<tr>
<td>G. Peter discovered John’s report on him.</td>
<td>5.2</td>
<td>8.4</td>
</tr>
<tr>
<td>H. John reported Peter’s criticisms of himself.</td>
<td>5.9</td>
<td>8.7</td>
</tr>
<tr>
<td>I. Peter saw John save him.</td>
<td>6.5</td>
<td>8.4</td>
</tr>
<tr>
<td>J. Peter saw John save himself.</td>
<td>4.3</td>
<td>6.5</td>
</tr>
</tbody>
</table>

(Adapted from Cook, 1990)

### 2.3.3 Intervention Studies

White (1995) conducted an intervention study to test the effects of instruction on the acquisition of English reflexives. Her subjects were 30 adult intermediate ESL learners. There were 19 Francophone and 11 Japanese ESL learners. There were three treatment groups. The first group (4 Japanese and 7 Francophone) received explicit positive grammar instruction on English reflexive pronouns. They were taught that English reflexives vary in morphological forms depending on number and gender, a reflexive can have a nonsubject NP as its antecedent under certain conditions. But they were not taught that long-distance binding is not allowed. The second treatment
group (6 Japanese and 5 Francophone) was a reading group in which subjects only read passages containing sentences with reflexives, and answered comprehension questions. The third treatment group (1 Japanese and 7 Francophone) was a control group. The treatment was 4 weeks long with three 20-minute sessions each week. Both an untimed grammaticality judgment task, i.e., asking learners to indicate whether a sentence is grammatical or not, and the truth-value judgment task, i.e., asking learners to answer a true/false comprehension question after reading a short passage, were used. They showed no significant differences among three groups. No effects of L1 were found. And almost no improvement was found from the pretest to the posttest for all groups. Because it was believed that with the framework of UG, some features were clustered and knowing one feature would automatically trigger the other. The study hypothesized that knowing that an object can be an antecedent would trigger the learning of the impossibility of long-distance binding English. No evidence was found to support the hypothesis that binding to objects triggered the learning that long-distance binding is prohibited. The grammaticality judgment task showed that subjects of all groups acquired the gender and number agreement of reflexives. This study seems to show that explicit instruction fails to improve the learning of binding principles. However, it is possible that negative evidence (i.e., informing subjects that long-distance binding is impossible for English) is necessary for subjects to reset the parameter. And different instruction methods, such as those using PI in the current study, may have different outcomes.
Another intervention study was conducted by White, Hirakawa and Kawasaki (1996). The major issue was whether teaching one supposedly related property of anaphora within the framework of UG can trigger the acquisition of the other properties without explicit instruction. For example, it has been proposed that the property of the possibility of long-distance binding of reflexives is related to the possibility of subject orientation, i.e., an antecedent must be a subject (Thomas, 1995). If one property is taught, can the other property be triggered? The target structure was the Japanese long-distance reflexive “zibun”. The question was whether teaching the property that “zibun” can take long-distance antecedents leads to the acquisition of the property that long-distance antecedents must be subject. There were 12 subjects from four different L1 backgrounds: 4 French, 2 English, 2 Korean, and 4 Chinese. Subjects were enrolled in a low intermediate level Japanese course. The training lasted 3 hours in total over a 4-week period, with 15-20 minutes per day, 2-3 times a week. The teaching provided exclusively positive evidence. Subjects were told that the antecedent of “zibun” can be outside of the clause but they were not taught whether the antecedent must be a subject or not. Training included reading and writing exercises. Results showed that more than half of the subjects knew that long-distance binding is allowed in Japanese. L1 backgrounds did not make a difference in the acquisition of the property. Two French speakers, one English, two Chinese and two Koreans acquired the long-distance property of “zibun”. Although the sample size in this study makes it hard to generalize, it does provide preliminary
evidence that instruction could help learners acquire the property of reflexives.

2.4 Linguistic Differences between English and Chinese Binding Principles

One relevant difference between English and Chinese is that English reflexives typically require a local antecedent whereas Chinese reflexives “ziji” can have either a long-distance antecedent or a local one. For example, in English “herself” only refers to “Flora” in “Mary thought that Flora blamed herself”. In the Chinese version of this sentence, “herself” can refer to either “Mary” or “Flora”. This means that a Chinese ESL learner must learn that long-distance binding is impossible in English. This is not easy to do because usually the binding constraints are not taught explicitly in the classroom or in standard textbooks. Thomas (1993) made the same observation.

Another difference is the notion of subject-orientation. Chinese requires the antecedent to be a subject whereas in English both subjects and objects can be antecedents of reflexive pronouns. So in “The doctor told the patient that Lisa praised herself”, in Chinese the patient cannot be the antecedent because it is an object. Also in English “Jill told Rose to perjure herself”, “herself” can refer to “Rose”. In Chinese, both “Rose” and “Jill” can be the antecedent.

To complicate the picture, another difference is that in Chinese there are two types of anaphora. One is the so-called morphologically simple (MS) or head (X0) reflexive, i.e., “ziji”, and the other is the morphologically complex (MC) or phrasal reflexive (XP) such as “taziji” (meaning: himself or herself). Both MS and MC can be
used alone, i.e., unlike the English “self”, “ziji” can be used independently. However, MC is usually used to emphasize the antecedent and takes only local antecedents (Cole and Sung, 1994, p.359). For example,

(8) Jack renwei Mike zhidaod John xihuan taziji.

Jack think Mike know John like himself

The above sentence means that Jack thinks that Mike knows that John likes himself. Here “taziji” can only refer to “John”. Researchers (e.g., Yuan, 1994) have tried to argue that this can transfer to the acquisition of local L2 binding. He argued that if learners linked the English “himself/herself” with the Chinese “taziji”, they could apply the Chinese rule in acquiring the binding constraint of the English reflexive. However, later research did not support this view (Yip and Tang, 1998; MacLaughlin, 1998). Results showed that learners’ judgments of the MC were not consistently locally bound. Furthermore, many learners who had acquired the English local binding principle allowed long-distance binding for the Chinese complex reflexive “taziji”. Another complication is that “taziji” is claimed to be ambiguous (Tang, 1985). It can be a local reflexive or, with the right information, a pronoun plus emphatic reflexive. When “taziji” is a pronoun plus emphatic reflexive, it can refer to a long-distance antecedent. Therefore in the above example, “taziji” can also refer to “Jack” or “Mike”.

Another difference is that in Chinese, pragmatics plays an important role in contrast to European languages such as English and French, in which the
interpretation of anaphora is governed by syntax. Huang (1994) has proposed a pragmatic theory of anaphora to account for the interpretation of anaphora in “pragmatic languages”. According to Huang, reflexive interpretation is governed by the I-principle, which means that it will find an antecedent that renders the most informative and stereotypical interpretation that is consistent with our general world knowledge. For example, in the Chinese version of “The little boy was angry that his father always hit himself”, according to Huang, “himself” is referentially dependent on the embedded subject “his father” and it is evaluated first. Although this is plausible, it is not the best interpretation because of our world knowledge that parents are more likely to hit children, not themselves and the little boy was angry. So “himself” is interpreted as referring to “the little boy”.

To sum up, English reflexives are locally bound while Chinese reflexives are not. The Chinese reflexive “ziji” is strictly subject-oriented. Pragmatics plays a relatively more important role in Chinese reflexive interpretation than in English.

2.5. Processing Instruction (PI)

The following sections introduce features of PI, its differences from traditional instruction, principles of PI, efficacy of PI, why PI is chosen as the instruction in the current study.
2.5.1 Features of Processing Instruction

According to VanPatten (2003), PI aims to alter learners’ inappropriate processing strategies that lead them to misinterpret or ignore forms in the input (See VanPatten [2002] for a detailed discussion). It is achieved by structured activities that make form-meaning connections salient to become intake, i.e., the part of input that is processed by learners. As discussed in the introduction, according to the Input Processing framework (VanPatten, 1996, 2003, 2004), learners derive only limited linguistic information from input and make certain form-meaning connections but not others. Research showed that learners establish form-meaning connections for content words first because they are more meaning-bearing and crucial for comprehension and communication. Learners, especially beginners, cannot afford to attend to all parts of a message because of a limited capacity of working memory. Further, the same information is often redundantly encoded in some other grammatical forms. For example, learners could ignore the past-tense marker –ed but still understand the message because usually they could get the same information from an adverb of time as in “Last week we visited Las Vegas”. According to VanPatten, those less meaning-bearing forms are only processed when learners are able to allocate attentional resources to attend to them.

There are three components in PI: explicit grammar explanation, referential activities, and affective activities. The latter two are called structured input activities. Explicit explanation includes explicit description of the target grammar and also alerts
learners about the wrong processing strategies that may prevent them from making form meaning mappings. For example, when learning Spanish, learners often misinterpret the first noun phrase as the subject of the sentence as in the following example:

(9) Lo ve Carlos.

Him See Carlos

Carlos sees him.

To help learners overcome the wrong strategy, PI will include an explicit explanation stating that in Spanish, the first noun phrase is not always the subject.

Referential activities are those listening and reading activities that force learners to process forms and structures to derive meaning, and they push learners away from incorrect processing strategies. Answers to referential activities are either right or wrong. One example of such activity in teaching Spanish grammar as in example 10 is that learners often listen to a recording and then choose “Who sees whom?”

Affective activities are those exercises in which learners express opinions and beliefs and involve themselves in the learning process. There are no right or wrong answers. Learners have to do something (e.g., by saying “Yes, that applied to me”.) after hearing or reading sentences. This activity will help learners take an active part in their learning process. It is emphasized that learners are not required to produce the target structure during the instruction.
2.5.2 Differences between Traditional Instruction and Processing Instruction

The following figure describes the basic process of L2 acquisition according to PI (Cited in Toth, 2006, p.322). PI posits that grammar instruction should start at the level of processing input, i.e., helping learners get intake from input. Again, intake refers to the part of input that learners process when trying to understand the message and hold in the working memory for further processing. So focused practice should happen in the process of “I” as labeled in the following figure.

I        II            III
input ————→ intake ————→ developing ————→ output

system

Focused practice

I= input processing; II=restructuring; III= access, production procedures

Figure 1: L2 acquisition processes according to PI

I        II            III
input ————→ intake ————→ developing ————→ output

system

Focused practice

I= input processing; II=restructuring; III= access, production procedures

Figure 2: Traditional practice
In contrast, traditional instruction follows the processes shown in Figure 2, which emphasizes the output exercises (Lee and VanPatten, 2003, p.133), regardless of the type of those exercises, i.e., whether they are mechanical or communicative drills. The PI model emphasizes the role of input. Processing of input feeds the developing system that can be accessed for output. Traditional practice is output-based, in which learners are first provided with grammar explanation and then are asked to practice by producing the target language. VanPatten argued that output practice only engages those processing mechanisms that are responsible for accessing the developing system but not those that are involved in feeding into the developing system. The output practice may develop fluency and accuracy in production but it may not help shape the grammar of the learners’ mental representations because, as shown in Figure 1, what feeds the developing system is intake, not output.

To sum up, traditional instruction focuses on output while PI focuses on input and is comprehension based. But PI is not just another comprehension pedagogy such as “input flood”, which teaches learners grammatical patterns with an excessive amount of language materials. And it takes processing mechanisms into learners’ processing of converting raw input to intake. It has several principles that are distinct. These will be discussed in the next section.

2.5.3 Principles of Processing Instruction

guidelines for constructing structured input activities (VanPatten, 2003, p.154).

1. One thing at a time.

2. Keep meaning in focus.

3. Learners must do something with the input.

4. Use both oral and written input.

5. Move from sentences to discourse.

6. Keep the learners’ processing strategies in mind. For example, when teaching learners the past tense marker –ed, activities should avoid letting learners use time adverbials such as “yesterday” and ignore the verb ending –ed to get the meaning.

Many replication studies (e.g., Salaberry, 1997; DeKeyser and Solkaski, 1996; Collentine, 1998; Nagata, 1998; Allen, 2000) that failed to replicate the original studies were criticized by VanPatten and colleagues on the grounds that they failed to design activities that conform to these guidelines and hence were not “genuine” PI. For example, Allen (2000) compared PI and traditional instruction and found that PI was not superior to traditional instruction. However, VanPatten (2002) pointed out that in their training materials of PI, learners only heard causative “faire” sentences such as “Jean fait promener le chien à Marie.” (John makes Mary walk the dog.) and did not hear the noncausative “faire” such as “Marc fait les valises pour Jean.” (Mark packs bags for John.). Thus learners could rely on some “mechanical” strategy, i.e., all sentences are causative since this is the grammatical point in which they are trained,
to answer the question. So their PI failed to comply with the sixth guideline.

2.5.4 Efficacy of Processing Instruction

VanPatten and Cadierno (1993) compared the effectiveness of PI and TI. The target structure was Spanish movement constructs containing OVS surface order preverbal direct object pronouns and participants were second-year Spanish students who were divided into three groups: a PI group (n=27), a TI group (n=26) and a control group (n=27). Learners of Spanish often misinterpret sentences such as “Lo ve Maria” as “He sees Mary” rather than “Mary sees him”. The PI group received structured input activities such as “select the picture that best goes with what you hear” and the TI group received paradigmatic grammar explanation of object pronouns (e.g., direct object pronouns such as me/te/la/nos/los/las are not the subject) followed by oral and written output practice including mechanical (i.e., those activities that do not need learners to attend to the meaning and usually there is only one correct answer), then meaningful (i.e., those drills that need learners to attend to the meaning of sentences to respond correctly but both parties involved in the activities already know the answer), and then communicative activities (i.e., those activities that need learners’ attention to meaning and the message contained in the response is unknown to the person who asks the question). Results found that on the sentence level interpretation test (e.g., learners hear a sentence and select a picture that corresponds to the sentence heard), PI significantly outperformed the other two
groups, while for the production test (as measured by the task “Complete the sentence based on the pictures you see”) both PI and TI groups made significant improvement and TI was as good as PI. The gains for the two groups were long-term effects as measured by 4-week delayed tests.

Since VanPatten and Cadierno (1993) published the original study on PI, many studies have been conducted and there are three major types of studies. One type of PI study has focused on comparing PI with the traditional instruction (TI), and this represents the earlier research (e.g., Cheng, 2002; Benati, 2001). The second type has focused on comparing PI with other types of instructions such as Meaning-based Output Instruction (MOI) (e.g., Morgan-Short and Bowden, 2006; Farley, 2001b). Another type of study has examined the differential effects of the three components of PI (e.g., Sanz and Morgan-Short, 2004). Also different structures and different L2s have been examined. For example, Benati (2005) examined English simple past-tense and compared PI, TI and MOI. Allen (2000) investigated French causative and compared PI and TI. In general, studies have confirmed the effectiveness of PI (Benati, 2001, 2005; Cadierno, 1995; Cheng, 2002; Farley, 2001a, 2001b; VanPatten and Sanz, 1995; VanPatten and Wong, 2004; Morgan-Short and Bowden, 2006). For example, R. Ellis (1999) reviewed the studies of PI and concluded that PI results in improved ability in comprehending targeted structures and it is superior to production-based instruction. In terms of production gains, PI is not better than production-based instruction but it is equally good and it seems to be more durable (R. Ellis, 1999).
However, the claim that PI is superior to TI has met challenges. Several studies (e.g., Morgan-Short and Bowden, 2006; Toth, 2006; DeKeyser and Sokalski, 1996; Salaberry, 1997; Collentine, 1998; Nagata, 1998; Allen, 2000; Erlam, 2003) have reported that TI or output practice is as good as PI. Although slightly different assessments and different treatments employed in those studies make it difficult to compare them to the original study operated under the PI framework, it does point out that more studies should be done to systematically evaluate the effectiveness of PI. One important improvement for future studies is that treatment must be faithful to the principles and guidelines of PI and other types of instruction must be matched to PI on all other variables other than those variables to be investigated (e.g., amount of input, length of instruction time). As discussed in section 2.5.3, the validity of those comparison studies is not without problems.

2.5.5 Rationale for Choosing PI as the Treatment

Processing instruction was chosen as the treatment for the current study for the following reasons:

First, it is a pedagogical technique derived from VanPatten’s theoretical model: input processing. It is one of the few pedagogical interventions that are based on psycholinguistic processes that learners experience when comprehending second language input. It takes the nature of real-time input processing and processing problems into L2 acquisition. VanPatten (2004, p.5-6) and Harrington (2004)
cautioned that the IP model is not an account of L2 sentence processing, especially not parsing. It intends to capture the conditions under which learners make form-meaning connections. It attempts to answer questions such as the developmental nature of form-meaning connections and factors that affect learners’ initial processing when they try to understand a sentence.

Second, PI has been extensively studied in the field of instructed second language acquisition (SLA) since the publication of VanPatten and Cadierno (1993). Many studies (e.g., DeKeyser and Sokalski, 1996; Marsden, 2006; Salaberry, 1997; Sanz and Morgan-Short, 2004; Takimoto, 2006; Toth, 2006; VanPatten and Sanz, 1995; VanPatten, 2002; VanPatten, 2005) have investigated the effectiveness of processing instruction. VanPatten (2002) claimed that PI is superior to traditional instruction (TI). TI refers to lessons consisting of explicit explanation of the target grammar and drills ranging from mechanical to communicative, to give students practice. However, PI is not without challenges. R. Ellis (1999) summarized the results of early PI studies and called for further research to clarify whether PI affects implicit knowledge or just raises noticing and understanding because research has not demonstrated that PI leads to gains in spontaneous use of target structures. Nearly all studies used only offline measures (such as interpretation tests in which participants listen to sentences and choose the corresponding pictures to represent what they hear, and production tests in which participants complete sentences using given words) for the assessment of the learning effects. Doughty (2004) also noticed this was a general problem of the field
of instructed SLA, i.e., offline measurements may just capture metalinguistic or conscious knowledge of the grammar. Psycholinguistic methods are ideal candidates to measure implicit knowledge. The use of explicit knowledge by learners in psycholinguistic tasks such as self-paced reading is minimized (Jiang, 2004, 2007). The current study will employ psycholinguistic tasks as assessment measures.

Third, PI has not been used to study English structures as often as Spanish structures (Morgan-Short and Bowden, 2006). It is important to extend the research to more languages and structures to test the technique.

2.6 Explicit Feedback versus Implicit Feedback in SLA

2.6.1 Feedback in SLA

Many studies have examined whether feedback is facilitative and what kinds of feedback benefit learners most (e.g., Lyddon, 2006; Rosa and Leow, 2004, R. Ellis, Loewen, and Erlam, 2006). These are important questions because if feedback is necessary for acquisition, teachers face a practical decision: to give feedback or not and what kind of feedback (explicit, recasts, negotiation moves such as clarification requests, confirmation checks and repetition). These questions are also important from a theoretical standpoint. For instance, it is important to know what role cognitive factors such as attention play in developing the interlanguage (Schmidt, 1995). There are different types of feedback. Negative feedback can be broadly defined as language data following an error made by a language learner. It may be facilitative for learners
to correct their errors on particularly difficult language forms (White, 1991). Negative feedback can be explicit or implicit (such as recasts [Spada, 1997]). Recasts are defined as the reformulation of an incorrect utterance following an error learners make. Ayoun (2001) used the computer program HyperCard to investigate the effects of negative and positive feedback on language acquisition. The target structures were two French past-tenses, *passé compose* and *imparfait*. There were three groups of participants: the recasting group who received implicit negative feedback, the modeling group who received pre-emptive positive evidence (i.e., the group received explicit grammar explanation before doing the activities) and the grammar group who got explicit positive evidence plus negative feedback. Results showed that the recasting group outperformed the grammar group, but not the modeling group.

Nicholas, Lightbown and Spada (2001), Long (2007) and Ellis and Sheen (2006) reviewed the literature on recasts and found that recasts can be beneficial to acquisition, especially when they are relatively explicit. However, it is unclear how effective recasts are in promoting what kind of learning (implicit vs. explicit) as argued by R. Ellis et al. (2006). As shown in Ellis and Sheen (2006), most studies take learner repair as evidence of learning, which is controversial because repair does not equal uptake (i.e., learners’ immediate response followed by teachers’ corrective feedback) and certainly not acquisition. Therefore the value of feedback may be overestimated.

Other studies point to an advantage of explicit feedback. Carroll and Swain
(1993) and Carroll (2001) found that the explicit feedback group in their study, which received metalinguistic information, outperformed other groups. They concluded that explicit feedback was most helpful in terms of improvement in language production. A recent study by R. Ellis, Loewen, and Erlam (2006) directly compared the effects of explicit feedback in the form of metalinguistic explanations and implicit feedback in the form of recasts. The target structure was the English past-tense –ed. Treatment was completed during two communicative tasks in which one group received recasts following any error in the target structure and the other group received metalinguistic explanation. For instance, if a learner in the recast group makes a mistake such as “I see a movie last week”, the teacher will respond “Oh, you saw a movie last week”. Instead, the teacher will respond “Here last week indicated a past action. The past tense of see is saw” to the same mistake to learners in the metalinguistic group. Measurements were oral imitation tests and GJ tests and a metalinguistic knowledge test. A posttest showed that the explicit group outperformed the implicit group for both the delayed (2 weeks) imitation and GJ tasks. Although the oral imitation task was designed to tap the implicit knowledge, it is far from an optimal measurement since explicit knowledge could still be deployed to perform the task. Some researchers (e.g., McDade, Simpson and Lamb, 1982) have already questioned the validity of imitation as a measure of implicit knowledge.

Rosa and Leow (2004) conducted a CALL study. They used computer-based tasks differing in the degree of explicitness by combining three variables: a pretask
providing explicit grammatical information; feedback concurrent to input processing; and different types (implicit or explicit) of feedback. Results showed that the explicit conditions were superior especially in production of both new and old items and in recognition of new items. Explicit feedback alone was effective on accuracy in the target structure. Other CALL studies have also reported the benefit of explicit feedback. Nagata (1993) also investigated the role of explicit and implicit written feedback. The traditional group received feedback without metalinguistic information while the other group received feedback containing metalinguistic explanation, i.e., explicit explanation of formal grammatical rules. The training was to write L2 sentences based on situations described in participants’ L1. Results showed that the group with metalinguistic information outperformed the other group for particles but not for verbal predicates and the former group outperformed the other group on a retention test. Nagata concluded that metalinguistic feedback was more effective if a target grammar is complex.

Contrary to the above studies, some studies have found that explicit feedback did not benefit learners more than other types of feedback. Recently, Lyddon (2007) investigated four types of corrective feedback (meaning-focused which ignores form errors as long as information is correct; implicit form-focused which gives feedback in the form of recast; non-metalinguistic explicit form-focused in which rejects unequivocally form errors without metalinguistic explanation (e.g., No, wrong answer); and metalinguistic which provides metalinguistic rules in addition to
unequivocal rejection of errors) and two types of enhancement (typographic enhanced input (e.g., by underlining or using bold fonts to draw learners’ attention to the target grammar; and unenhanced input) in the acquisition of French prepositions a/au/en/aux in his dissertation. The materials were digitized and participants interacted with a computer program during the experiments. Analyses showed that all groups improved in the target structure and no significant difference was found among the 8 treatment groups. What exactly makes the results in this study different from other studies is unclear. It could be that depending on different target structures, whether they are communicatively redundant or not, feedback has different effects. In this study the target form is of less communicative value in PI’s terms because learners can get the gist of the message without attending to the form. However, this study is not the only one that points to the null advantage of explicit feedback. The following studies also showed that explicit feedback is not necessary for gains in L2 acquisition.

2.6.2 Feedback in PI

Traditional PI studies did not specify the role of feedback in treatment. Therefore, it is likely that different studies provided learners with different kinds of feedback, since most PI treatment was administered by a language instructor. More recent studies have investigated this issue. Sanz and Morgan-Short (2004) conducted a computer-assisted study to examine the role of explicit feedback in PI. Materials were essentially the same as VanPatten and Cadierno (1993) and VanPatten and Sanz (1995)
but were digitized to make them suitable for computer-delivered lessons. The target structure was Spanish preverbal direct object pronouns. Spanish has a flexible word order and it is common for L2 learners to misinterpret sentences such as “Lo besa la chica” (The girl kisses him) as “He kisses the girl” (VanPatten, 1984). This may be due to the so-called “first noun strategy”, i.e., learners assign the first NP as the agent of the preverbal noun (Slobin and Bever, 1982). There were two variables, i.e., [+/-] explicit explanation and [+/-] feedback. Therefore there were 4 groups with different combinations of the two variables. All participants received the same structured input and all got feedback that told them whether their answer was correct. The group with explicit explanation received explicit information about the structure and processing strategies. For example, this group was told that:

“In Spanish a sentence may have a subject and an object. An object is generally defined as a thing or person on which an action or process is performed. Thus, in the sentence John writes letters, John is the subject and letters is the object.... What is the subject of the verb in the following sentence?...”

(Sanz and Morgan-Short, 2004, p.56)

The group with feedback received information telling them whether they were right and why if they were wrong. Results showed that there were no differences among the groups as assessed by one interpretation task and two production tasks (one was a sentence completion task, the other was a video-retelling task). All groups
improved significantly. This suggests that the structured input activities contribute to the major gains in PI. Neither explicit information nor explicit feedback seemed essential. However, one drawback is that only accuracy was measured. Even VanPatten (2002) himself called for further study using reaction time measures to further test the effects of PI.

Sanz (2004) examined the role of implicit and explicit feedback in PI. First- and second-year Spanish learners participated and the target structure was Spanish preverbal direct object pronoun constructions. There were two experiment groups, i.e., the implicit group and the explicit group. For the former, 16 participants received structured input activities and yes/no feedback. For example, participants were instructed to select a picture that matches a sentence they heard and the sentence contains the target grammar, i.e., Spanish clitic pronouns. Participants in this group were told whether their selection was correct. For the explicit feedback group, 12 participants received structured input activities and explicit feedback including the reason why the answer was wrong, i.e., by informing them of the wrong processing strategies. For example, they were told that in Spanish a direct object pronoun is not the subject and they should not use the first-noun-strategy, i.e., assigning the first noun phrase as the agent of the verb. Testing involved one sentence interpretation task and two production tasks, one sentence completion, the other video-retelling. Results showed that the structured input activities alone led to improved gains in ability to interpret and produce the targeted structure. Feedback had no differential effects,
whether it was implicit or explicit. This finding is contradictory to those who advocate providing explicit CALL feedback. The studies on feedback in SLA in general and feedback in CALL suggest that explicit feedback results in better performance in language acquisition. Contrary to this conclusion, the work from PI suggests that feedback is not necessary in the development of interlanguage if the activities are manipulated and structured as in PI. The effectiveness of explicit feedback depends on the timing of the feedback, whether it is available during production or during input decoding, or on the conditions, whether it is combined with more input exposure or whether activities are structured as those in PI. Because most of the studies used only offline measures, which may tap into more explicit knowledge than implicit knowledge, it remains to be tested which type of feedback is most effective for the acquisition of processing routines in L2.

2.7 Conclusion

Chapter 2 reviewed important previous studies on L1 online pronominal processing, L2 anaphora acquisition studies, linguistic differences between Chinese and English binding principles, previous research in PI and the role of feedback. It laid the theoretical foundation for the current study. The next chapter will introduce the current study which will integrate all those elements discussed above and investigate those issues.
CHAPTER 3: THE CURRENT STUDY

3.1 Overview of the Current Study

The study intends to use psycholinguistic methods to examine L2 learners’ online processing of anaphora and to find out what strategies or cues they use and to what extent they can process coreference as L1 learners do. It also intends to alter learners’ inappropriate processing strategies by processing instruction. It will investigate the role of feedback in CALL in the acquisition of anaphora. There are three experiments in the study. The first two experiments will find out what ESL learners do when they comprehend anaphora online and the third is an intervention study. It will test whether processing instruction can help ESL learners acquire the necessary processing strategies.

3.2 Research Questions and Hypotheses

There are four research questions:

1. Can L2 learners process anaphora as L1 learners? Are they sensitive to the structure constraint and the feature constraints? Specifically, can they employ those cues to understand anaphora during online sentence processing?

1.1 Featural constraints, i.e., the role of features of potential antecedents such as gender and number on the interpretation of anaphora. Will these features be applied automatically?
1.2 Implicit causality: verb cues. As discussed in the literature review, verbs can be biased toward NP1 or NP2 as their antecedents. Can L2 speakers use verb cues automatically?

1.3 Structural constraints, i.e., the role of structural constraints on coreference processing. Can L2 learners use structural information in antecedent assignment? I will limit my inquiry to the acquisition of Binding Principle A.

2. Is L2 competence (as measured by grammaticality judgment) reflected in L2 performance (as measured by online comprehension methods such as self-paced reading)?

3. Can we alter L2 learners’ processing strategies by computer-delivered processing instruction? I will limit my focus on reflexives.

4. What is the role of explicit feedback (metalinguistic, i.e., no and why wrong vs. implicit, i.e., yes/no feedback) in CALL (Following Sanz, 2004)?

The Hypotheses are:

1. Advanced L2 learners will be sensitive to featural cues (gender and number) and verb cues. They may not be sensitive to structural constraints and employ inefficient processing strategies. For example, due to the interference of their L1, they may use their L1 strategy and allow long-distance binding. Thus, their acquisition of coreference processing constraints may be selective.

2. Their competence will lag behind performance. Learners often know the rules such as adding –ed to a verb to form a past tense, but make mistakes when speaking
3. PI will be effective in the sense that it will improve participants’ scores as measured by the paper-pencil written test. PI will be ineffective to alter L2 learners’ inefficient processing as measured by online methods.

4. Explicit feedback will benefit participants more than implicit feedback in the CALL program.

The rationale for each of the above hypotheses will be discussed in the following chapters for each experiment. Experiment 1 will answer research questions 1.1 and 1.2. Experiment 2 will answer research question 1.3. Together they will answer research question 2. Experiment 3 will answer research questions 3 and 4.

## 3.3 Methodology

The following sections will provide a general discussion of the methods used in the study. They will include a comparison of online and offline methods and introduce some of the standard psycholinguistics methods in SLA. Finally, they will also discuss computer-delivered processing instruction.

### 3.3.1 Online Methods versus Offline Methods

Testing methodologies used in the field of SLA can be roughly divided into two types based on whether they are sensitive to timing issues. Those methods which are not time sensitive, i.e., cannot reveal the moment-by-moment behavior of participants,
and require an untimed judgment of some kind, are offline methods. Traditional methods, such as untimed grammaticality judgment and sentence-picture matching, belong to this category. In contrast, time-sensitive methods, such as moving window self-paced reading tasks, are considered “online” because they provide data about processing difficulty as a sentence unfolds.

3.3.1.1 Traditional Method: Grammaticality Judgment

Grammaticality judgment (GJ) is one of the most widely used methods in SLA. In this task, participants are asked to judge the acceptability of sentences based on their well-formedness. For example, researchers working within the framework of Universal Grammar (UG) have used judgment data to test whether UG is accessible to L2 learners. It is argued that the task can be used for non-naturally occurring sentences created by experimenters and reveals participants’ internalized knowledge, i.e., competence, which is different from performance, i.e., what learners actually do with language. Despite the popularity of the task, its validity and reliability as a pure reflection of a person’s grammar have been questioned. For example, Birdsong (1989) pointed out that GJ merely taps metalinguistic performance and may not be direct evidence of linguistic competence. Besides linguistic knowledge, other factors such as response biases (e.g., a given subject may tend to reject most sentences), the nature of different structures, and inter and intra learner differences can affect such judgments. Researchers have found it difficult to determine what leads to learners’ judgments (R.
Ellis, 1990, 1991). Some learners used semantic interpretation, some applied implicit knowledge and some relied on explicit knowledge. This concern has something to do with the relationship between metalinguistic judgments and grammatical knowledge. Judgments are introspective in nature and may belong to a general nonlinguistic cognitive system and thus GJs would not tap into grammatical competence (Bever, 1970). The fact that different learners use different strategies to determine grammaticality indicates that GJs lack internal consistency.

Researchers also found that the task may elicit variable judgments. For example, R. Ellis (1990), using a test and retest paradigm, found that learners changed their judgments 22.5% of the time for an advanced group, and 45% of the time for the intermediate group. In a test-retest experiment, Han (2000) also reported that learners changed 36% of their judgments between the same tasks for the same structures.

One of the potential problems of GJ is that learners could apply explicit knowledge and thus the results tell us little about how an L2 learner will use the language in the real world and little about their real-time processing strategies. Although timed or speeded GJ tasks could reduce the effect of learners’ reliance on explicit knowledge, it is still possible that learners’ explicit knowledge is involved (DeKeyser, 2003). To avoid this drawback, psycholinguistic methods will be employed in the current study. It will give us a better sense of how an L2 learner is going to use a second language in the real world.
3.3.1.2 Online Method: Self-Paced Moving Window Reading Tasks

Psycholinguistic methods such as reading time tasks have been widely used to study sentence comprehension in native speakers (e.g., Ferreira and Clifton, 1986; Frazier and Clifton, 1998; Trueswell, 1996, among many others), but far less so to study L2 comprehension. As discussed above, offline methods such as GJ only inform us about the end product of sentence comprehension processes and can tell us little about what learners actually do while they understand sentences.

In a typical self-paced moving window reading task, participants sit in front of a computer monitor and read sentences word-by-word. At any given point, participants press a key to advance to the next word as soon as they are ready to continue reading. The current word disappears when the next word appears on the screen. Participants integrate the meaning of each word into the sentence incrementally. Participants cannot backtrack. To make sure participants read the sentences for general comprehension, it is common that a comprehension question follows each sentence or randomly sized group of sentences. The computer records participants’ reaction time to each word and each response to the comprehension questions. Feedback can be provided to inform participants whether the response to the comprehension question is correct.

One major assumption behind the self-paced reading task is that reading time reflects reading difficulty. Relatively longer reading time (compared to a baseline) reflects processing difficulty. For example, it has been demonstrated that, compared to
grammatical sentences, ungrammatical sentences take longer to process for both L1 readers (e.g., Pearlmutter, Garnsey, and Bock, 1999; Segalowitz, 2003) and L2 readers (for certain structures; Jiang, 2004). For L2 learners, the task has been adopted to investigate whether learners have integrated their grammatical knowledge into their processing systems. For example, in Jiang’s study (2004), participants read sentences like the following (the region of interest is underlined here, not in the experiment):

3.1a. The bridges to the island were about ten miles away.

3.1b. *The bridge to the island were about ten miles away.

Results showed that native speakers slowed down at both “were” and the “about” when reading 3.1b, the ungrammatical version. This indicates that readers noticed the number disagreement in the sentence. Nonnative speakers, however, did not take longer to read these words, indicating that they were not sensitive to the failure of agreement. In the same study, L2 learners were sensitive to violations of verb subcategorization. Although it has been shown that L2 learners are sensitive to some grammatical violations such as those involving verb subcategorization, they appear to be insensitive to others, such as those involving the plural –s morpheme. Overall, few types of ungrammaticality have been tested with online methods. The current study will begin to fill this gap in L2 research by using the moving window reading paradigm to examine the processing of anaphora.
3.3.2 Computer-Delivered Processing Instruction

Experiment 3 is an intervention study using computer-delivered PI as the method of instruction. There are several advantages to computer-presented PI. First, traditional classroom instruction involves teachers. This may introduce confounding variables because teachers may teach differently to different students, even for the same lesson. The computer program, instead, delivers the same stimuli to all participants in the same way. Second, in a classroom setting, “teacher talk” and “student talk” can both serve as input to learners, making it difficult to control the amount of input each learner obtains in the classroom. With a computer program, each learner gets the same amount of input. Third, in addition to accuracy data, the use of computer technology in the study permits response times to be recorded so that time on task may be analyzed. Studies have shown that, in general, time on task is positively correlated to learning (e.g., Bloom, 1974). Nikolova (2002) reported that for second language vocabulary learning, time on task was a variable that could reverse the research findings for the study. When time was not considered, participants who were not given annotations for the new words in a text, but were asked to create annotations for the target words during the treatment phase, learned vocabulary significantly better than participants who were provided annotations and were not instructed to create annotations themselves. When time on the task was taken into account, no difference was found between the two groups in terms of vocabulary learning. Fourth, it is easier to replicate a study involving computerized instruction
than teacher-delivered instruction. The use of computers allows control of individual and environmental variables such as teacher differences, the timing of stimuli presentation and timing of feedback. The type of feedback could be provided as needed, at different timing intervals. Learners could also be assigned to treatment groups randomly, which is often difficult if administered in a classroom. Fifth, because PI requires no production of language, the computer-delivered lesson is particularly appropriate for this instruction; learners will simply interact with the computer via a mouse. In the classroom, learners are more likely to produce the target structure and could impact the results because learners’ output can be other learners’ input. Besides the above advantages as an experimental technique, it also has advantages for learners. Training is individualized since learners control the pace of the lesson. Stimuli may be a combination of audio, visual and text, which is potentially more interesting than in-class lessons.

3.4 Conclusions

This chapter provides an introduction to the current study, including the research questions and hypotheses, and the rationale of the methods to be used. It has been shown that online methods such as the self-paced moving window reading task, which provides data about language processing over the course of a sentence, reflect how well L2 learners process the second language as it is presented, which is, for the most part, what they will be doing outside the classroom. Also, it was argued that, for
research purposes, a computer-assisted intervention has distinct advantages over the more standard classroom instruction.

The next three chapters will describe the actual three experiments, with the first two dealing with online processing strategies of anaphora and the third with computer-delivered PI research.
CHAPTER 4: EXPERIMENT 1

4.1 Introduction

Experiment 1 was designed to answer research question 1, in particular questions 1.1 and 1.2, i.e., whether second language learners automatically use featural constraints and verb cues to assign antecedents when they interpret pronouns and reflexives online. It is an important experiment because the results will also inform the feasibility and design of Experiment 2 with respect to the featural cues that can be manipulated to test learners’ sensitivity to structural cues. The following sections will describe the research questions and hypotheses, methodology and instrumentation, participants, procedures, data analysis, and finally results and discussion of the first experiment.

4.2 Research Questions and Hypotheses

The purpose of Experiment 1 is to test L2 learners’ sensitivity to featural constraints on coreference, and so-called “implicit causality” associated with verbs. Can advanced L2 learners of English use gender and number information and verb cues in online processing? If they can, they should automatically slow down when they read sentences whose ungrammaticality involves these types of information.

It is assumed that native speakers (NSs) will, of course, show such slowdowns in all cases. It is predicted that nonnative speakers (NNSs) will be sensitive to gender
and verb cues, but they will not be sensitive to number cues. These predictions follow from previous work, including research by Jiang (e.g., 2007) which found that advanced L2 learners had difficulty detecting the anomalous sentences containing the plural morpheme –s during their online reading of texts. NNSs showed no reading time differences in the underlined four positions of sentences such as “The visitor took several of the rare coins in the cabinet” and “The visitor took several of the rare coin in the cabinet” (Jiang, 2007). Native speakers slowed down after the word “coin” (compared to “coins”), presumably reflecting the difficulty of reading the ungrammatical sequence “several of the coin”. With respect to gender and verb information, there is some indication that NNSs may be sensitive to this. For example, Felser, Roberts, Gross, and Marinis (2003) found that L2 learners are sensitive to lexical-semantic information. Results showed that advanced L2 learners of English used the lexical-semantic properties of the preposition (such as of versus with) to resolve relative clause attachment ambiguities in sentences such as “The dean like the secretary of the professor who was reading a letter” versus “The dean liked the professor with the secretary who was reading a letter”. Their preference for the attachment was similar to English native speakers, i.e., they preferred the first noun phrase “the secretary” as the person who was reading a letter for the first sentence and preferred the second noun phrase “the secretary” as the person who was doing the action. Therefore, it is hypothesized that advanced English learners will show sensitivity to those lexical cues and thus will take longer to read sentences containing
gender mismatches and misleading verb cues.

4.3 Methodology

As discussed in Chapter 3, the self-paced reading methodology will be used. This is a standard psycholinguistic method in which subjects press a key to advance word-by-word throughout a sentence. It is assumed that a slowdown in reading an anomalous sequence reflects participants’ sensitivity to the grammatical constraints under investigation. A comparison of reading time at the critical positions in grammatical vs. ungrammatical sentences, underlined in the upcoming example, will be used to determine whether L2 learners are sensitive to gender, number, and verb cues. For example, given the two sentences below,

The careless pedestrian found themselves covered with mud.

The careless pedestrian found himself covered with mud.

The former takes longer to read starting at the reflexive because it is ungrammatical. As discussed above, this technique has been used extensively with native-language readers and successfully with L2 learners (e.g., Jiang, 2004, 2007).

In addition to the online reading task, traditional “paper and pencil” tests and background questionnaires will be administered to assess grammatical knowledge and overall proficiency in English. This is necessary both to ensure homogeneity among the subjects, and to allow comparisons with other studies in the literature.

To sum up, the research design is mixed, with both quantitative and qualitative
aspects. The quantitative study measures participants’ sensitivity to specific cues while the qualitative part gathers data about learners’ language learning background.

4.4 Instrumentation

This section introduces the test materials. All instruments were specially created for the experiment. The nature and purpose of each instrument are given below.

4.4.1 Self-Paced Reading Tasks

Reflexive Sentences: Simple sentences containing a reflexive were created, along with their ungrammatical counterparts, as in the examples below. A complete list of the items can be found in Appendix A.

Example stimuli for the antecedent-reflexive number agreement cues:

The careless pedestrians found themselves/himself covered with mud.

Were the pedestrians covered with water?

The famous actress prepared themselves/herself to face the crowd.

Example stimuli for the gender agreement cues:

The lonely grandfather made himself/herself a cup of tea.

Did the grandfather cook some soup?

The hungry waitress ordered herself/himself a big burger.

Example stimuli for the implicit causality (verb) cues:

The mother amused the father because he/she told funny jokes at dinner.
(NP1-biased sentences)

The boy admired the girl because he/she was so intelligent. (NP2-biased sentences)

Was the boy admired?

Some of the materials were adapted from Osterhout and Mobley (1995) (See Appendix A). For the “number” sentences, half the sentences contained a plural subject and half contained a singular subject. Likewise, for the “gender” sentences, half contained a feminine subject and half contained a masculine subject. In order to ensure the sentences are understandable to the participants, only those judged to be easy to read by the researcher are included in the test items. This is also confirmed by participants’ informal reports after the test.

“Implicit Causality” sentences: The stimuli were adapted from a number of published studies (e.g., Rigalleau and Caplan, 2000; Stewart et. al., 2000; McDonald and MacWhinney, 1995). For each sentence, there are two clauses connected by the word “because” to indicate a causal relation. To repeat one example introduced previously, in “Jane admires Jill because she is reliable”, readers link “she” with “Jill” (the second noun phrase, NP2) instead of “Jane” (the first noun phrase, NP1) because the verb “admire” takes an “experiencer” subject, which is not a typical causer of the action. In contrast, “Jane amused Jill because she told funny jokes”, readers link “she” with “Jane” (NP1) instead of “Jill” (NP2) because the object of the verb “amuse” is usually an experiencer, and again, experiencers are not typical causal agents. An
additional factor in pronoun resolution is a reader preference for “parallel function”: a subject will be favored as the antecedent of a subject pronoun. This could add to the NP1 bias and detract from the NP2 bias. For clarity, the latter will be called “NP1-biased sentences” and the former will be called “NP2-biased sentences”.

Two counter-balanced lists were constructed. Each list contained 16 grammatical “number” sentences, 16 ungrammatical “number” sentences, 16 grammatical “gender” sentences and 16 ungrammatical “gender” sentences, and 32 “implicit causality” sentences, half that resolve in favor of the bias and half against the bias. So the total number of critical sentences was 96, with 16 tokens per type. No subjects saw all variants of a sentence. Half of the sentences were followed by yes/no comprehension questions; these were included to ensure that participants were actually reading for general comprehension. Most of the comprehension questions followed grammatical sentences; however, to keep participants paying attention to both grammatical and ungrammatical sentences, several ungrammatical sentences were followed by a comprehension probe. For these, either a Yes or a No answer was treated as correct. There were also 32 filler sentences.

**4.4.2 Offline Written Test**

Materials also included a written test administered to NNSs to assess their explicit knowledge of gender, number constraints and “implicit causality” constraints in the resolution of anaphora.
Because the focus of this study is on the online processing of anaphora, it is crucial to establish that participants at least have explicit knowledge of the constraints under study. It is unlikely (though not impossible) that they will automatically process this information online if they do not demonstrate knowledge of these constraints in untimed tasks. There were 15 items altogether. There were 3 items targeting the masculine “himself”, 3 items targeting “herself”, 3 items targeting a number-matched reflexive, 3 targeting NP1-biased verbs and 3 NP2-biased verbs. Items were similar or identical to those used in the online self-paced reading task. Items were presented in pseudo-random order. A complete list of items appears in Appendix B. The following presents the sample stimuli for each category:

Sample item testing the gender cue:

The lonely grandfather made_____ (himself/herself) a cup of coffee.

The gracious wife introduced_____ (himself/herself) to the guests.

Sample item testing the number cue:

The careless pedestrians found _____ (themselves/himself) covered with mud.

Sample item testing the NP1-biased verb cue:

The mother amused the father because_____ (he/she) told funny jokes.

Sample item testing the NP2-biased verb cue:

The widow envied the bachelor because _____ (he/she) won the national lottery.
4.4.3 Language Profile

ESL participants’ language background information was collected via questionnaires (See Appendix C). The questionnaire was adapted from Jiang (2007). It was a one-page survey with questions about participants’ demographic information about age, age at which a participant began to learn English, length of residence in the U.S., their language learning experience and their self-ratings of English proficiency in listening, reading, speaking and writing skills. Such information is informative with respect to both homogeneity of the subject group and comparability with subjects tested in previously published research.

4.5 Participants

Two groups of participants were tested: English native speakers and advanced ESL learners. Forty-one native English speakers served as the control group. Nine were excluded from the analysis because of high error rates in the written test or in the reading comprehension questions task (more than 20%). All of them were undergraduate students at the University of Arizona. Most of them were registered in the psychology subject pool and were taking introductory psychology courses. Some of them were taking introductory linguistics classes. They received course credit for participating in the experiment.

Forty-one Chinese ESL learners participated in the study. Seven participants were excluded from the analysis due to high error rates in the written test or in the reading
comprehension questions (more than 20%). They were graduate students at the University of Arizona and their spouses. All participants had normal vision or were corrected to normal vision. They all spoke Chinese as their first language. Most of them were science major graduate students. NNSs’ ages ranged from 19 to 41, with an average of 30. Based on their TOEFL score and self-ratings of their English proficiency, they were considered “advanced” ESL learners. Their average TOEFL score was 627, with a range of 560 to 677. The minimum requirement of TOEFL for admission to most universities in the U.S. is 550. Their average self-ratings of English in the four skills, i.e., speaking, listening, reading and writing were, respectively, 3.9, 4.1, 4.2, 3.9 (on a scale of 1 to 5 with 5 being most proficient). Participants were well-educated, with an average of formal education of 20 years. They had been in the U.S. for an average of 3.7 years and they had been studying in English-speaking countries for an average of more than 3 years. Most participants started to learn English after puberty. Only one person started to learn English at the age of 3. Her TOEFL score was 623. The following table summarizes the background information of the nonnative participants in Experiment 1:
### Table 3: Nonnative speakers’ language profile in Experiment 1

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Minimum</th>
<th>Maximum</th>
<th>SD</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>30</td>
<td>19</td>
<td>41</td>
<td>4.9</td>
<td>34</td>
</tr>
<tr>
<td>TOEFL scores</td>
<td>624</td>
<td>560</td>
<td>677</td>
<td>27.7</td>
<td>30</td>
</tr>
<tr>
<td>Age starting English</td>
<td>12</td>
<td>3</td>
<td>16</td>
<td>2.4</td>
<td>34</td>
</tr>
<tr>
<td>Years of formal education</td>
<td>20</td>
<td>13</td>
<td>25</td>
<td>2.6</td>
<td>34</td>
</tr>
<tr>
<td>Years of residence in the U.S.</td>
<td>3.7</td>
<td>1</td>
<td>10</td>
<td>2.4</td>
<td>34</td>
</tr>
<tr>
<td>Years of education in</td>
<td>3.3</td>
<td>1</td>
<td>9</td>
<td>1.8</td>
<td>34</td>
</tr>
<tr>
<td>English-speaking countries</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-rating of English proficiency</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Speaking</td>
<td>3.9</td>
<td>3</td>
<td>5</td>
<td>0.7</td>
<td>34</td>
</tr>
<tr>
<td>Listening</td>
<td>4.1</td>
<td>3</td>
<td>5</td>
<td>0.7</td>
<td>34</td>
</tr>
<tr>
<td>Reading</td>
<td>4.2</td>
<td>3</td>
<td>5</td>
<td>0.6</td>
<td>34</td>
</tr>
<tr>
<td>Writing</td>
<td>3.9</td>
<td>2</td>
<td>5</td>
<td>0.8</td>
<td>34</td>
</tr>
</tbody>
</table>

### 4.6 Procedures

Participants were tested individually in a quiet test booth. After providing signed consent, participants were seated in front of a computer monitor and button-box for the self-paced moving window reading task. They were told that they would read sentences for general comprehension one word at a time and as soon as they read each word they would press the right-hand button to advance to the next word. The current word disappeared when the next word appeared. They were told to proceed as quickly as they could, but to make sure that they fully understood what they were reading. They were also told that occasionally there would be a yes/no comprehension question about what they had just read and they were expected to press either the right-hand button for “Yes” or the left-hand button for “No”. They were told to read as quickly as possible and answer comprehension questions as accurately as possible.
To familiarize the participants with the task, they read 10 practice items of various types. DMDX, an experimental software developed by Forster and Forster at the University of Arizona (See Forster and Forster, 2003), was used to present the materials and record the participant’s reaction time to each word in a sentence, along with responses to each comprehension question. The self-paced reading task took less than 35 minutes.

After the online task, participants filled out the language profile questionnaire. The questionnaire took less than 10 minutes to complete. In order to minimize the possibility of sensitizing participants to the focus of the study, it was decided that the paper-pencil written tests should follow the online task in Experiment 2.

4.7 Data Analysis and Results

The online reading data were analyzed only for participants who performed the task satisfactorily. Exclusion of participants was based on several criteria. First, participants’ comprehension rates were calculated. Those who had an error rate higher than 20% were excluded from further analysis. Nine native speakers were excluded for this reason, leaving the total number of native speakers at 32. Those native speakers’ error rates ranged from 2.5% to 20%, with an average error rate of 10%. Five nonnative speakers were excluded, leaving the total number of nonnative speakers at 36. Those nonnative speakers’ error rates fell between 3.8% to 20%, with an average error rate of 14%. The comprehension rates suggested that the two groups
of participants were following the instructions to read the sentences for general comprehension and that they understood the sentences well.

Second, for nonnative speakers, test scores in the written test were calculated. The written test was administered immediately after Experiment 2 was completed. Those whose error rates were more than 10% were excluded from the analysis. Two nonnative speakers were excluded for this reason, leaving the total number of nonnative speakers at 34. The nonnative speakers’ error rates fell between 0% to 10%, with an average error rate of 7%. The written test results showed that nonnative speakers in the study have mastered the structures investigated in the study as measured by the offline task.

In addition, following standard procedure, reaction times (RTs) that were two SDs longer or shorter than the same participant’s mean, or higher or lower than the high and low cutoffs set at 2,000 and 200 ms, respectively, were excluded. These data trimming procedures, along with missing data and display errors, accounted for 5.4% of the data. Of the remaining data, thirty two mean RTs were computed for each participant, one for each test position (there are four test positions) in each condition (there are two conditions: grammatical vs. ungrammatical) for each cue (there are four cues: the gender cue, the number cue, the NP1-biased verb cue and the NP2-biased verb cue). These above calculated means were subjected to the statistical program SPSS for the participant analyses. For the item analyses, sixteen means were calculated for each of the ninety six sentences, one for each of the four test positions
in each of the four conditions. The participants’ RTs for each test position, each condition, and each cue are given in the following tables. Data of NSs and NNSs are given below separately.

Table 4: Native speakers mean RTs (ms) at four test positions of an example sentence for grammatical and ungrammatical sentences involving number cues and gender cues

<table>
<thead>
<tr>
<th>Cue</th>
<th>Number</th>
<th>Gender</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test position</td>
<td>1 found</td>
<td>2 self</td>
</tr>
<tr>
<td>Grammatical</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>449</td>
<td>436</td>
</tr>
<tr>
<td>Ungrammatical</td>
<td>438</td>
<td>436</td>
</tr>
<tr>
<td>Difference</td>
<td>11</td>
<td>0</td>
</tr>
</tbody>
</table>

**significant at .05 in both subject and item analyses.

Table 5: Native speakers mean RTs at four test positions in example sentences for plausible and implausible sentences involving verb cues

<table>
<thead>
<tr>
<th>Cue</th>
<th>NP1-biased verb</th>
<th>NP2-biased verb</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test position</td>
<td>1 because</td>
<td>2 s/he</td>
</tr>
<tr>
<td>Plausible</td>
<td>410</td>
<td>402</td>
</tr>
<tr>
<td>implausible</td>
<td>414</td>
<td>402</td>
</tr>
<tr>
<td>Difference</td>
<td>-4</td>
<td>0</td>
</tr>
</tbody>
</table>

*significant at .05 in subject analysis only.

**significant at .05 in both subject and item analyses.
Table 6: Nonnative speakers mean RTs (ms) at four test positions for grammatical and ungrammatical sentences involving number cues and gender cues

<table>
<thead>
<tr>
<th>Cue</th>
<th>Number</th>
<th>Gender</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test position</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1 found</td>
<td>2 self</td>
</tr>
<tr>
<td>Grammatical</td>
<td>560</td>
<td>550</td>
</tr>
<tr>
<td>Ungrammatical</td>
<td>556</td>
<td>553</td>
</tr>
<tr>
<td>Difference</td>
<td>4</td>
<td>-3</td>
</tr>
</tbody>
</table>

*significant at .05 in subject analysis only.
**significant at .05 in both subject and item analyses.

Table 7: Nonnative speakers mean RTs at four test positions for plausible and implausible sentences involving verb cues

<table>
<thead>
<tr>
<th>Cue</th>
<th>NP1-biased verb</th>
<th>NP2-biased verb</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test position</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1 because</td>
<td>2 s/he</td>
</tr>
<tr>
<td>Grammatical</td>
<td>507</td>
<td>444</td>
</tr>
<tr>
<td>Ungrammatical</td>
<td>518</td>
<td>462</td>
</tr>
<tr>
<td>Difference</td>
<td>-11</td>
<td>-18**</td>
</tr>
</tbody>
</table>

*significant at .05 in subject analysis only.
**significant at .05 in both subject and item analyses.

The data for NNSs and NSs were analyzed separately. The SPSS Paired-Samples T-Test was chosen to compare the participants’ mean RTs for grammatical/plausible and ungrammatical/implausible sentences at four different positions and for four cues. For each group, thirty-two pairs were compared both for subject analysis and item analysis. Because the prediction is directional, i.e., ungrammatical/implausible versions will take longer than grammatical/plausible ones, the statistical test is
one-tailed. The detailed results of the analyses can be found in Appendix D.

Since we expect no differences in RTs in positions 1 and 2 and expect some differences in RTs in positions 3 and 4, the following discussion will deal with positions 1 and 2 and then positions 3 and 4.

4.7.1 Reading Time Patterns at Positions 1 and 2

As can be seen in the above tables 4, 5, 6 and 7, the participants’ RTs at the first two positions showed no reliable differences, with the exceptions of the first position for the number cue sentences by NSs, the second position for the gender cue sentences by NSs, the NP2-biased verb sentences by NSs, and the NP1-biased verb sentences by NNSs. However, the 11 ms difference for the first exception was significant only in the by-item analysis but not significant in the by-subject analysis. There is no clear explanation for this and it will not be discussed further. The 15 ms difference for the second exception was significant in both the subject and item analysis, which suggests that participants noticed the error immediately. For the third exception, the difference was only significant by subjects and will not be discussed further. For the fourth exception, the RT difference was significant in both the subject analysis and the item analysis. This pattern indicates that participants were very sensitive to the NP1 verb cues and they slowed down even at the pronoun when it did not match the gender of the favored NP. One of the well-attested effects of self-paced reading tasks is the spillover effect which doesn’t manifest itself at the position where
the mismatch occurs but appears in the next region. This is apparent in the current experiment in which, in most cases, participants slow down one or two words after they encounter the mismatch. Therefore, the results at the first two positions overall confirmed the no-difference predictions for both groups of participants and the cues tested.

4.7.2 Reading Time Patterns at Positions 3 and 4

Native speakers: NSs and NNSs’ RTs at the third and fourth positions showed different patterns for different cues tested. For NSs, RTs between the grammatical and ungrammatical sentences at the two positions for the number cue were statistically significant both in subject analysis, $t_1(31)=4.8$, $p<.05$ for position 3, $t_1(31)=4.1$, $p<.05$ for position 4 and in item analysis, $t_2(31)=5.7$, $p<.05$ for position 3 and $t_2(31)=3.6$, $p<.05$ for position 4. The same was true for the gender cue sentences, $t_1(31)=7.5$, $p<.05$ for position 3, $t_1(31)=6.3$, $p<.05$ for position 4 and in item analysis, $t_2(31)=6.3$, $p<.05$ for position 3, $t_2(31)=6.0$, $p<.05$ for position 4. For the NP1-biased verb cues, the same was true. The RTs’ difference between grammatical and ungrammatical sentences in position 3 was significant in both subject analysis, $t_1(31)=1.9$, $p<.05$ and in item analysis, $t_2(31)=1.8$, $p<.05$. In position 4, results were statistically significant in both subject analysis, $t_1(31)=2.6$, $p<.05$ and in item analysis, $t_2(31)=2.1$, $p<.05$. For the NP2-biased verb cues, a slightly different pattern was found for native speakers. Nevertheless, it still fits in the expectations. The RTs’
difference between grammatical and ungrammatical sentences in position 3 was found to be statistically different in both subject analysis, \( t_1(31) = 3.2, \ p < .05 \), and in item analysis, \( t_2(31) = 3.4, \ p < .05 \). In position 4, the difference in RTs was found to be statistically nonsignificant in both subject analysis, \( t_1(31) = 1.3, \ p = .200 \), and in item analysis, \( t_2(31) = 1.1, \ p = .271 \).

**Nonnative speakers:** For NNSs, the patterns were similar to those of the NSs with the exception of the pattern for the number cues. For the gender cue, RTs between the grammatical and ungrammatical sentences at the two positions were statistically significant both in subject analysis, \( t_1(33) = 5.6, \ p < .05 \), for position 3, \( t_1(33) = 3.3, \ p < .05 \) for position 4, and in item analysis, \( t_2(31) = 4.4, \ p < .05 \) for position 3 and \( t_2(31) = 2.2, \ p < .05 \) for position 4. For the number cue, unlike the NSs, NNSs showed no reliable differences in both subject analysis, \( t_1(33) = 1.6, \ p = .056 \), and in item analysis, \( t_2(31) = 1.2, \ p = .114 \), for the third position. For the fourth position, it was found that results were significant in subject analysis, \( t_1(33) = 2.1, \ p < .05 \), but nonsignificant in item analysis, \( t_2(31) = .90, \ p = .189 \). For the NP1-biased verb cue, in position 3, the 30-ms difference was significant in subject analysis, \( t_1(33) = 2.8, \ p < .05 \), and approached significance in item analysis, \( t_2(31) = 1.5, \ p = .077 \). In position 4, the 18-ms difference was significant in subject analysis, \( t_1(33) = 1.9, \ p < .05 \), but was nonsignificant in item analysis, \( t_2(31) = 1.1, \ p = .135 \). For the NP2-biased verb cue, in position 3, the difference was significant in subject analysis, \( t_1(33) = 2.1, \ p < .05 \), but nonsignificant in item analysis, \( t_2(31) = 1.6, \ p = .059 \). In position 4, the difference was
significant in both subject analysis, \(t_1(33)=3.1, p<.05\), and in item analysis, \(t_2(31)=1.9, p<.05\).

### 4.8 Discussion and Conclusion

Experiment 1 was designed to answer the following research questions 1.1 and 1.2: Can L2 learners process pronouns and reflexives as L1 speakers do? Are they sensitive to the featural constraints on coreference? Specifically, can they use featural cues automatically to constrain coreference resolution during online sentence processing?

1.1 Featural constraints, i.e., the role of features of potential antecedents such as gender, number on the interpretation of anaphora. Will these features be accessed and used automatically?

1.2 Implicit causality: Verbs are biased toward NP1 or NP2 as their antecedents (See section 2.2.3 for examples). Can L2 speakers use verb cues automatically?

The results show that native speakers automatically use all the cues available in the sentences to assign antecedents and when there is a mismatch between the cues, they encounter reading difficulties. Nonnative speakers’ data patterned in the same direction; however, not all cues yielded reliable differences in RTs. The L2 speakers were sensitive only to certain cues, such as the gender cue, and their sensitivity to other cues such as the number cue was very weak. Therefore, the answer to the research questions 1.1 and 1.2 are affirmative, but only partially. Even advanced ESL learners can only use certain cues to assign antecedents during their online reading.
They automatically use the gender cue and the verb cues, but their automatic use of the number cue was not as strong as other cues.

This is in line with Jiang (2007), who also found that even advanced ESL learners did not process the English morpheme plural –s the same as L1 speakers do. Using a self-paced reading task, Jiang (2007) asked both NSs and NNSs of English to read grammatical and their ungrammatical counterparts involving the plural –s morpheme or verb subcategorization. Example sentences are:

The visitor took several of the rare **coins** in the cabinet.

*The visitor took several of the rare **coin** in the cabinet.

The teacher wanted the **student** to start all over again.

*The teacher insisted the **student** to start all over again.

A different pattern was found for NSs and NNSs. NSs slowed down one or two words after the point of ungrammaticality. While NNSs slowed down only for those sentences containing verb subcategorization violations, suggesting that they were ignoring the plural –s morpheme during their online processing or were not able to keep track of the word for a plural NP. Several researchers have found that the acquisition of the plural morpheme tends to remain nonnative-like (Long, 2003; Schmidt, 1983; White, 2003). L2 learners, even after a 20-year period of exposure in the target language, still did not use correct plural forms of nouns in their spoken English. It has been suggested that ESL learners selectively integrate grammatical knowledge, or their “automatic competence” is selective. “Selective automatic
competence” can also describe the results of Experiment 1. Assuming that automatic competence can be measured in terms of subjects’ sensitivity to grammatical errors during the online reading, the current results support the idea that ESL learners’ automatic competence is selective. They can use the gender cue highly automatically, indicating that they have integrated the knowledge of how to access and use the gender information in the processor. Their failure to automatically use the number cue for antecedent assignment suggests that they have not integrated knowledge about using the number information in the processor. Can ESL learners develop the ability to use the plurality information in their language processor? Given the fact that data for the NNSs in this experiment are toward the expected direction, it remains to be tested if even more advanced ESL learners are sensitive to the cues under investigation.

To sum up, Experiment 1 established that advanced ESL learners were sensitive to certain cues to pronoun and reflexive resolution such as the gender and the verb information, but not the number feature. The next research question is whether ESL learners use the structural information to assign antecedents.
5.1 Introduction

Experiment 1 found that advanced ESL learners were more sensitive to certain cues than to others during their online language processing. The next question is whether they are sensitive to structural information. Experiment 2 explores this question. As discussed in the introduction, there are structural differences in anaphora in Chinese and English. Chinese allows both local and long-distance binding for reflexives while English allows only local binding. For example, in the Chinese equivalent of “Mike believes that Tom trusts himself”, “himself” can refer to either “Mike” or “Tom”. Offline studies (e.g., Thomas, 1989) have shown that Chinese learners of English can apply the English rule. But we do not know whether they apply it online, since explicit rule knowledge may come into play during offline tasks. Because the goal of language learning is to be able to use the language automatically, it is important to investigate the online language processing of binding constructions.

5.2 Research Questions and Hypotheses

Can advanced L2 learners (L1 Chinese, L2 English) process English reflexives in the same manner as native speakers of English? Will they apply the Binding Principle A online? Will they transfer their L1 strategy for understanding reflexives to L2? To what extent will they be influenced by pragmatic knowledge? These are the research questions to be answered in Experiment 2.
It is hypothesized that for the offline test, ESL learners will be able to make the right judgment and show that they have mastered the English constraint. For the self-paced reading study, reading time on or just after the reflexive for L2 readers will not be the same as for L1 readers. The L1 group will be sensitive to the binding principle. Grammatical sentences will read faster than the ungrammatical counterparts. The reading patterns will not be influenced by pragmatics. As research has indicated that L2 readers underuse grammatical information during online sentence parsing (e.g., Jiang, 2004), CELs will not be sensitive to the binding principle. And their L1 strategy transfers to L2 processing. This means that they allow long-distance binding and use lexical information such as gender information to select antecedents. If gender information can single out one antecedent, sentences are easier to process. Sentences are difficult to read if both noun phrases in the sentence can be the antecedent when using the L1 strategy. Another interesting question is to what extent L2 readers will be influenced by pragmatics. Compared to neutral sentences, biased sentences are more likely to induce the transfer of L1 strategies, i.e., allowing long-distance binding when reading English sentences. So it is more likely to find L1 transfer among those sentences. This will mean that they employ inappropriate processing strategies. This will show that they have not acquired the Binding Principle A, i.e., they have not integrated into the processor that reflexives are locally bound and long-distance binding is not possible.
5.3 Methodology

The methods are basically the same as in Experiment 1. Both an offline comprehension test and a word-by-word moving window self-paced reading task were administered. Both native speakers and ESL learners were tested to make sure the materials test what they were designed to test.

5.4 Instrumentation

The following introduces the design of the materials for Experiment 2. All materials were specially designed for the experiment.

5.4.1 Self-Paced Reading Tasks

Since Experiment 1 found that both native and nonnative speakers were highly sensitive to the gender cue when understanding reflexives during their online reading, the gender cue was manipulated in Experiment 2 to test whether participants are sensitive to the structural cue, i.e., whether they apply Binding Principle A during their online reading.

Four counter-balanced lists were constructed. Each list contained 12 sentences of each type as shown in the following sample stimuli, i.e., Type 1a-4a and Type 1b-4b. There were two large categories, i.e., pragmatically neutral sentences and pragmatically biased sentences. For the neutral sentences, the pragmatic meaning of the sentence doesn’t bias toward the choice of the antecedent. For example, in “The son remembered that the father introduced himself at the meeting”, the pragmatic
meaning of the sentence does not favor either of the two noun phrases “the son” or “the father”. Both are plausible. For the biased sentences, the sentence meaning makes the selection of the antecedent structurally incorrect in English. For example, in “The little girl was happy that the mother bought herself a nice toy last week”, the sentence meaning favors the interpretation that “herself” means “the little girl” since we know that mothers usually buy toys for their girls and thus the little girl felt happy.

Of course, according to English grammar, here “herself” corefers to the mother. The sentence construction was modeled after Demirci (2000). Some of the sentences were adapted from previous studies (e.g., Demirci, 2000; Osterhout and Mobley, 1995; Sturt, 2003) and the rest were constructed by the researcher (See Appendix E). One novice Chinese ESL learner was consulted to see if the sentences rendered the intended effect or confusion, i.e., was asked to interpret the sentences and see if both antecedents were possible or if one was better than the other. One expert in psycholinguistics made necessary changes to make the sentences sound as smooth and natural as possible.

Within each category, there were four types of sentences based on whether the gender cue matches the gender of a potential antecedent and whether the noun phrase or the potential antecedent is accessible as an antecedent according to Binding Principle A. For example, in the following sample sentences, Type 1a is called a “both match” type because both the noun phrase “the son” and “the father” are masculine and the gender matches the reflexive “himself”. Of course, according to the binding
principle, only the second noun phrase “the father” is an accessible antecedent of the reflexive. Other types of sentences are: “only NP2 match” as shown in Type 2a, “only NP1 match” as shown in Type 3a, and “no match” as shown in Type 4a. The noun phrases were half gender specific terms such as father and mother and half common proper names such as Mary and David. Gender neutral terms, such as the chair, were avoided as much as possible. For each category, half sentences used male noun phrases to start the sentences and half used female noun phrases to begin the sentences. No subjects saw all variants of a sentence. Half of the sentences were followed by yes/no comprehension questions. Participants were not probed explicitly about the reflexives in the task so that they were not sensitized to the structure tested explicitly. It was decided that questions should follow some ungrammatical sentences so that participants would not develop a strategy to ignore all ungrammatical sentences. There were 32 filler sentences, serving as distractors.

Sample stimuli:

Pragmatically neutral sentences:

The son remembered that the father/mother introduced himself/herself at the meeting.

-Was the son introduced at the meeting?

Type 1a: accessible-match/inaccessible match (both match): The son remembered that the father introduced himself at the meeting.

Type 2a: accessible-match/ inaccessible mismatch (only NP2 match): The son
remembered that the mother introduced herself at the meeting.

Type 3a: accessible mismatch/inaccessible match (only NP1 match): The son remembered that the mother introduced himself at the meeting.

Type 4a: accessible mismatch/inaccessible mismatch (no match): The son remembered that the father introduced herself at the meeting.

Pragmatically biased sentences:

The little girl was happy that the father/mother bought herself/himself a nice toy last week.

+Was the little girl happy?

Type 1b: accessible-match/inaccessible match (both match): The little girl was happy that the mother bought herself a nice toy last week.

Type 2b: accessible-match/inaccessible mismatch (only NP2 match): The little girl was happy that the father bought himself a nice toy last week.

Type 3b: accessible mismatch/inaccessible match (only NP1 match): The little girl was happy that the father bought herself a nice toy last week.

Type 4b: accessible mismatch/inaccessible mismatch (no match): The little girl was happy that the mother bought himself a nice toy last week.

5.4.2 Offline Written Test

For CELs, a written test was administered (See Appendix F) to establish that participants knew the rules offline. Items were similar or identical to the online study.
There were 12 items in total, testing 3 different sentence categories: neutral sentences, NP1-biased sentences, NP2-biased sentences. There were also 3 distractors to discourage participants from using strategies such as always choosing the nearest local noun phrase as the antecedent. Sentences were pseudo-randomized in the actual test. The following sentences are examples for each of the categories:

Neutral sentences:

Rose said that Jill voted for herself.
Can “herself” refer to Rose? Yes No
Can “herself” refer to Jill? Yes No
Can “herself” refer to somebody else? Yes No
NP1-biased sentences:

The hungry girl was happy that the mother bought herself an ice cream.
Can “herself” refer to the hungry girl? Yes No
Can “herself” refer to the mother? Yes No
Can “herself” refer to somebody else? Yes No
NP2-biased sentences:

Jill heard that the famous actress talked about herself on TV.
Can “herself” refer to Jill? Yes No
Can “herself” refer to the famous actress? Yes No
Can “herself” refer to somebody else? Yes No
Distractors:

The man with the little boy bought himself a new house in the city.
Can “himself” refer to the man? Yes No
Can “himself” refer to the little boy? Yes No
Can “himself” refer to somebody else? Yes No

5.4.3 Language Profile

The same language profile was used as in Experiment 1 (Appendix C). ESL learners’ background information was collected through the profile.

5.5 Participants

Two groups of participants took part in the experiment. Fifty native English speakers served as the control group. They were from the same population as in Experiment 1: They were undergraduate students enrolled in introductory psychology or linguistic courses at a southwestern university. They received partial course credit for participating in the experiment. The NNSs group was the same as in Experiment 1, except for two NNSs who did not participate in the second experiment. Therefore, in total, thirty-nine NNSs completed the experiment. However, only sixteen of the participants were included in the following data analysis. The others were excluded due to their high error rate on the results in the paper-and-pencil test (more than 20%), which confirms the premise that the binding principle is hard to master. Even advanced learners still allow long-distance binding in the written test. For the participants who mastered the structure, their overall proficiency in English is “advanced”. As shown in the following, their TOEFL scores were high (M=628, SD=32.1) and their language proficiency self-ratings (on a scale of 1 to 5 with 5 the
best) also showed that they were advanced ESL learners. The following presents the background information of the NNSs in Experiment 2:

Table 8: Nonnative speakers’ language profile in Experiment 2

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Minimum</th>
<th>Maximum</th>
<th>SD</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>29.6</td>
<td>19</td>
<td>36</td>
<td>4.8</td>
<td>16</td>
</tr>
<tr>
<td>TOEFL scores</td>
<td>628</td>
<td>560</td>
<td>677</td>
<td>32.1</td>
<td>15</td>
</tr>
<tr>
<td>Age starting English</td>
<td>11</td>
<td>3</td>
<td>16</td>
<td>3.0</td>
<td>16</td>
</tr>
<tr>
<td>Years of formal education</td>
<td>20</td>
<td>13</td>
<td>25</td>
<td>3.1</td>
<td>16</td>
</tr>
<tr>
<td>Years of residence in the U.S.</td>
<td>3.4</td>
<td>1</td>
<td>8</td>
<td>1.9</td>
<td>16</td>
</tr>
<tr>
<td>Years of education in English-speaking countries</td>
<td>3.3</td>
<td>1</td>
<td>7</td>
<td>1.4</td>
<td>16</td>
</tr>
<tr>
<td>Self-rating of English proficiency</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Speaking</td>
<td>4.1</td>
<td>3</td>
<td>5</td>
<td>0.9</td>
<td>16</td>
</tr>
<tr>
<td>Listening</td>
<td>4.4</td>
<td>3</td>
<td>5</td>
<td>0.6</td>
<td>16</td>
</tr>
<tr>
<td>Reading</td>
<td>4.4</td>
<td>3</td>
<td>5</td>
<td>0.7</td>
<td>16</td>
</tr>
<tr>
<td>Writing</td>
<td>4.0</td>
<td>3</td>
<td>5</td>
<td>0.9</td>
<td>16</td>
</tr>
</tbody>
</table>

5.6 Procedures

Procedures in Experiment 2 were the same as in Experiment 1. Participants were tested individually in the same psycholinguistic lab. After giving written consent, participants completed the self-paced reading task on a computer in a small test booth. The instructions were essentially the same as in Experiment 1. They were told to respond as quickly and as accurately as possible. Experiments 1 and 2 were at least one week apart to minimize a carryover effect. Participants practiced ten items to get familiar with the task. The experiment used DMDX (See Forster and Forster, 2003) to present the stimuli and record participants’ reaction times. The task took less than 35
minutes for all participants. Participants could take three breaks during the experiment, which was different from Experiment 1, in which participants finished the experiment without breaks. Breaks were added in Experiment 2 based on participants’ feedback of Experiment 1 to give participants opportunities to relax. After finishing the online reading task, participants spent 10 to 18 minutes taking the paper-and-pencil written test for both Experiments 1 and 2.

5.7 Data Analysis and Results

Before examining participants’ online data, their scores on the written test were calculated. Those who had an error rate higher than 10% were excluded from further analysis. Sixteen NNSs were excluded for this reason, leaving the total number of NNSs at 22. The written test showed that the binding principle under study is very difficult for ESL learners. Most Chinese ESL learners allow both local binding and long-distance binding in English as they allow them in Chinese.

Then participants’ comprehension rates were calculated based on their responses to the comprehension questions. Those whose error rates were higher than 20% were excluded from further analysis. The native speakers’ error rates ranged from 5% to 20%, with an average error rate of 10%. Ten native speakers were excluded for this reason, leaving the total number of NSs at 40. Five NNSs were excluded for this reason, leaving the total number of NNSs at 16. The nonnative speakers’ error rates were from 6% to 20%, with an average error rate of 13%. The comprehension rates
suggested that participants understood the sentences well.

The next step was to examine participants’ RTs in the four critical positions for various sentence types. Data trimming procedures were the same as in Experiment 1, i.e., those RTs that were 2 SDs longer or shorter than the same participant’s mean, or higher or lower than the high and low cutoffs set at 2,000 and 200 ms, respectively, were excluded. These trimming procedures, along with missing data and display errors, accounted for 5.8% of the data. Of the remaining data, 32 mean RTs were computed for each participant, one for each test position (four test positions) in each condition (four conditions: both match, only NP2 match, only NP1 match and no match) for each sentence category (the neutral sentences and the pragmatically biased sentences). These calculated means were subjected to the statistical program SPSS for participant analyses. Because the prediction was directional, i.e., grammatical sentences will be read faster than the ungrammatical versions, the statistical test was one-tailed. For the item analyses, sixteen means were calculated, one for each of the four test positions in each of the four conditions. The participants’ RTs for each test position, each condition, and each sentence category are given in the following tables. Data of native speakers and NNSs are presented below respectively.
Table 9: Native speakers’ mean RTs (ms) at four test positions for neutral sentences and pragmatically biased sentences involving the structural cue for conditions “both match” and “no match”

<table>
<thead>
<tr>
<th>Test position</th>
<th>Neutral</th>
<th>Biased</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>introduced</td>
<td>bought</td>
<td>self</td>
</tr>
<tr>
<td>Both match</td>
<td>420</td>
<td>444</td>
</tr>
<tr>
<td>No match</td>
<td>414</td>
<td>444</td>
</tr>
<tr>
<td>Difference</td>
<td>6</td>
<td>0</td>
</tr>
</tbody>
</table>

**significant at .05 in both subject and item analyses.

Table 10: Native speakers’ mean RTs (ms) at four test positions for neutral sentences and pragmatically biased sentences involving the structural cue for conditions “only NP1 match” and “only NP2 match”

<table>
<thead>
<tr>
<th>Test position</th>
<th>Neutral</th>
<th>Biased</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>introduced</td>
<td>bought</td>
<td>self</td>
</tr>
<tr>
<td>Only NP1 match</td>
<td>419</td>
<td>433</td>
</tr>
<tr>
<td>Only NP2 match</td>
<td>419</td>
<td>435</td>
</tr>
<tr>
<td>Difference</td>
<td>0</td>
<td>-2</td>
</tr>
</tbody>
</table>

**significant at .05 in both subject and item analyses.
Table 11: Nonnative speakers’ mean RTs (ms) at four test positions for neutral sentences and pragmatically biased sentences involving the structural cue for conditions “both match” and “no match”

<table>
<thead>
<tr>
<th>Test position</th>
<th>Neutral</th>
<th>Biased</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 introduced</td>
<td>2 self</td>
</tr>
<tr>
<td>Both match</td>
<td>494</td>
<td>517</td>
</tr>
<tr>
<td>No match</td>
<td>501</td>
<td>536</td>
</tr>
<tr>
<td>Difference</td>
<td>-7</td>
<td>-19</td>
</tr>
</tbody>
</table>

**significant at .05 in both subject and item analyses.

Table 12: Nonnative speakers’ mean RTs (ms) at four test positions for neutral sentences and pragmatically biased sentences involving the structural cue for conditions “only NP1 match” and “only NP2 match”

<table>
<thead>
<tr>
<th>Test position</th>
<th>Neutral</th>
<th>Biased</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 introduced</td>
<td>2 self</td>
</tr>
<tr>
<td>Only NP1 match</td>
<td>510</td>
<td>533</td>
</tr>
<tr>
<td>Only NP2 match</td>
<td>497</td>
<td>513</td>
</tr>
<tr>
<td>Difference</td>
<td>13</td>
<td>20</td>
</tr>
</tbody>
</table>

*significant at .05 in subject analysis only.
**significant at .05 in both subject and item analyses.

The RT data for NNSs and native speakers were analyzed via the SPSS program.

Paired-Samples T-Tests were used to compare participants’ mean RTs for four conditions at four different positions and for two sentence categories. For each group, sixteen pairs were compared for subject analysis and item analysis, respectively.
Results can be found in Appendix G.

5.7.1 Reading Time Patterns at Positions 1 and 2

As can be seen from the above tables, the participants’ RTs at the first two positions showed no reliable differences for both NSs and NNSs for the four different sentence categories. This was expected for the first position since all words in the sentences were essentially the same up to that position. For position 2, RTs could be the same due to the well-attested spill-over effect of self-paced reading as discussed in Experiment 1. So the results at the first two positions confirmed the no-difference predictions for both groups of participants and the structural cues tested.

5.7.2 Reading Time Patterns at Positions 3 and 4

Native speakers: NSs and NNSs’ RTs at the third and fourth positions showed similar patterns for different conditions tested. For NSs, RTs between the “both match” and “no match” sentences at the two positions for the neutral sentences were statistically significant both in subject analysis, t1(39)=6.4, p<.05 for position 3, t1(39)=2.2, p<.05 for position 4 and in item analysis, t2(47)=4.9, p<.05 for position 3 and t2(47)=2.0, p<.05 for position 4. The same was true for the comparison between “only NP1” and “only NP2” sentences, t1(39)=3.3, p<.05 for position 3, t1(39)=1.9, p<.05 for position 4, and in item analysis, t2(47)=2.2, p<.05 for position 3, t2(47)=2.9, p<.05 for position 4.
For the biased sentences, a similar pattern was found for native speakers. The RTs’ difference between “both match” and “no match” sentences in position 3 was found statistically significant both in subject analysis, $t_1(39)=-3.7, p<.05$ for position 3, $t_1(39)=-2.7, p<.05$ for position 4, and in item analysis, $t_2(47)=-4.2, p<.05$ for position 3, and $t_2(47)=-2.1, p<.05$ for position 4. A slightly different pattern was found for “only NP1 match” sentences and “only NP2 match” sentences. In position 3, differences in RTs were significant in both subject analysis, $t_1(39)=4.9, p<.05$, and in item analysis, $t_2(47)=4.5, p<.05$. In position 4, the difference in RTs was not found to be statistically significant in either subject analysis, $t_1(39)=1.2, p=.119$, or in item analysis, $t_2(47)=1.3, p=.104$. Although the 11-ms RT difference in position 4 was not significant, it was in the same direction as in position 3.

**Nonnative speakers:** For NNSs, the patterns for the biased sentences were the same as those of the NSs while the patterns for the neutral sentences were only slightly different. For the biased sentences, RTs between “both match” and “no match” sentences at the two positions were statistically significant both in subject analysis, $t_1(15)=-2.8, p<.05$ for position 3, $t_1(15)=-2.0, p<.05$ for position 4, and in item analysis, $t_2(47)=-2.5, p<.05$ for position 3, and $t_2(47)=-2.0, p<.05$ for position 4. For the comparison between “only NP1 match” sentences and “only NP2 match” sentences, NNSs yielded the same pattern as NSs. In position 3, the difference was significant in both subject analysis, $t_1(15)=2.0, p<.05$, and in item analysis, $t_2(47)=1.6, p<.05$. In position 4, however, it was found that the 11-ms difference was
nonsignificant in both subject analysis, \( t_1(15) = .89, p = .194 \), and in item analysis, \( t_2(47) = .98, p = .166 \).

For the neutral sentences, NNSs showed reliable differences in some but not all expected positions. Nevertheless, in general, the data were consistent with expectations. For the third position, it was found that results were nonsignificant in both subject analysis, \( t_1(15) = -1.3, p = .115 \), and in item analysis, \( t_2(47) = -1.1, p = .150 \). The RT difference was in the right direction, i.e., participants slowed down 25 ms.

The RT difference between “both match” and “no match” sentences at position 4 were statistically significant both in the subject analysis, \( t_1(15) = -1.7, p < .05 \), and in the item analysis, \( t_2(47) = -2.0, p < .05 \). When the RT difference in position 3 between “only NP1 match” sentences and “only NP2 match” sentences was compared, it was found that results were significant both in subject analysis, \( t_1(15) = 2.8, p < .05 \), and in item analysis, \( t_2(47) = 2.5, p < .05 \). In position 4, the results were significant in subject analysis, \( t_1(15) = 2.5, p < .05 \), and approached significance in item analysis, \( t_2(47) = 1.6, p = .06 \).

To sum up, it is expected that no RT differences in positions 1 and 2 will be found and there should be differences in positions 3 and 4 if participants are sensitive to the structural cue. The results in general confirmed the expectations. For NSs, they slowed down within two positions after the position where the ungrammatical word occurs, in the online reading task. They showed no sign of allowing long-distance binding online. The results confirmed the predication that NSs are highly sensitive to
the structural constraints on binding for both neutral sentences and pragmatically biased sentences. NNSs were also sensitive to the structural cue. They showed the same pattern as NSs when reading pragmatically biased sentences and a similar pattern when reading neutral sentences.

5.8 Discussion and Conclusion

The goal of Experiment 2 was to test whether ESL learners use structural constraints automatically to interpret anaphora during their online sentence comprehension. Specifically, the questions were: Will they apply Binding Principle A during online reading? Will they transfer their L1 constraints to L2? To what extent will their processing of anaphora be influenced by pragmatic knowledge?

NSs showed reliable differences in RTs within two words after encountering the ungrammatical reflexive for both neutral sentences and pragmatically biased sentences. This replicated the standard finding in self-paced reading tasks that participants will slow down one or two words after reading the ungrammatical word. Results also suggested that NSs automatically use the structural cues to assign antecedents, and when there is a feature mismatch between an anaphor and its antecedent, they encounter reading difficulties. The combined results from Experiment 1 and Experiment 2 suggest that NSs use both featural, semantic cues and structural cues in the sentences during their online processing of anaphora.

Do NNSs apply Binding Principle A during online reading? For those NNSs who
have acquired the principle as measured by the paper-and-pencil test, online data showed that they were sensitive to the structural cues and they applied Binding Principle A. When the only noun phrase that matched the gender of the reflexive was inaccessible according to Principle A in English (but accessible in their L1), participants did not apply the Chinese principle. This can be best seen from the comparison in RT difference between “only NP1 match” sentences and “only NP2 match” sentences. Data showed that the former type of sentences took longer to read than the latter type of sentences. If participants transferred the Chinese principle of allowing long-distance binding, the former type of sentences would have been considered as grammatical and posed no reading difficulty. The results suggested that advanced ESL learners did not transfer their L1 strategy to online L2 processing.

Unlike the findings in Demirci (2000), pragmatics did not pose great difficulties to those advanced ESL learners included in Experiment 2 during their online reading tasks. Even when the sentence meaning favored the antecedent that is accessible in the sentence structure for the anaphora if they apply their L1 strategy, they resisted the temptation and still took much longer to read those sentences when compared to the grammatical counterparts. This suggested that they have successfully acquired the processing strategy in their L2 English. L2 learners can achieve automatic processing both semantically (i.e., use semantic information during online reading) and structurally (i.e., use structural information during online reading). This finding runs counter to recent claims made by Clahsen and Felser (2006) who claim that L2
learners only use semantic information and compute shallower structure during their online processing. In Experiment 1, it was found that participants used semantic information such as verb information and also featural information such as gender information to assign antecedents. This is consistent with the findings that L2 learners use semantic or lexical information during online processing. In Experiment 2, it was found that advanced L2 learners also employed structural information to assign antecedents. This is not consistent with the claim that L2 learners’ processing is “shallow” since advanced ESL learners do use hierarchically structural information in the current study. Without analyzing the structure of the input, it is unlikely that ESL learners could have applied Binding Principle A in this study. The binding principle is highly abstract and many consider it part of UG (Chomsky, 1981). So the structure should not be considered as shallow. The natural conclusion is that advanced ESL learners do use deep structural information for online processing.

The written test showed that the binding principle posed difficulty to many of the participants. More than half of the participants scored under 90% on the test. This is perhaps not surprising given that the binding principles are not explicitly taught in class. This is true both according to the researcher’s own experiences and participants’ self-reports after the experiment. This is one of the reasons that this structure is an ideal testing target because the test results reflect natural language acquisition, i.e., acquiring language in a natural environment without explicit instruction. In addition, there is lack of sufficient evidence in the input that informs learners of the rule.
Context or other cues such as gender congruence usually limit the choice of antecedents. Participants have few opportunities to encounter those sentences containing conflicting cues such as semantic cues and structural cues as used in the current experiments. Participants’ informal reports confirmed this observation.

To sum up, Experiment 2 found that advanced ESL learners used structural information when assigning antecedents online. This is an exciting finding since previous studies have failed to find evidence for automatic use of structural information in online processing. Experiment 2 also revealed that many ESL learners have not fully acquired the constraints on anaphora, especially Binding Principle A. The difficulty of this structure led to the next experiment: What can teachers do to help learners develop ability in automatic processing and develop their fluency. Experiment 3 will explore this question.
CHAPTER 6: EXPERIMENT 3

6.1 Introduction

Experiments 1 and 2 found that advanced ESL learners process anaphora in a similar way as NSs. Offline studies also found that many ESL learners have not acquired the full properties of anaphora, especially the binding principles. They are not sensitive to the structural information. So what can teachers do to help them? White (1995) showed that the traditional instruction did not help. One possible way is through processing instruction (PI). As discussed in the literature review, PI is an input-based focus-on-form pedagogical intervention which takes psycholinguistic processing into consideration. The challenge is to design valid structured activities for the targeted grammar. To better control intervening variables, “Director” was chosen as the software to deliver the instruction via computer-assisted language learning techniques. The following sections will introduce the specific research questions and methodology and results for Experiment 3.

6.2 Research Questions and Hypotheses

The major purpose of Experiment 3 is to test whether PI leads to interlanguage development as it claims. More specifically, can we alter L2 learners’ processing strategies by computer-delivered processing instruction? Teaching one thing at a time is one of the principles of PI, so the forms will be limited to the reflexives “himself”
and “herself” in this experiment. The strategy to be changed is: from over-relying on lexical cues such as gender to relying on structural information. Another research question is: What is the role of feedback (metalinguistic, i.e., no and why wrong vs. implicit, i.e., yes/no feedback)? As discussed in Chapter 2, previous studies yielded inconclusive findings to this question and called for further research.

It is hypothesized that PI will improve participants’ interpretation and production scores significantly as measured by the offline written tests. However, their automatic coreference processing will not be improved. One reason is that what they have acquired is explicit metalinguistic knowledge, which may not be deployed during automatic processing. It is also hypothesized that the types of feedback will make a difference in the outcome of the results. Previous studies (e.g., Lyddon, 2006) found that manipulations of explicit and other types of feedback did not affect the acquisition of French, “aux”, “au”, and “a la”. However, explicit feedback could potentially make a difference in the participants’ linguistic behaviors. First, the feedback could raise participants’ awareness of the structure and provide explicit rules about the structure. Second, it is provided only when needed and it is tailored to each individual’s learning needs. Third, it informs participants about the wrong strategies that they are using and this conforms to the PI’s principle, i.e., focusing on the incorrect use of the strategy. On the contrary, implicit feedback only informs participants whether the answer is right or wrong. They will have to guess the rules themselves with the possibility of failing to do so. Therefore, difference will be
expected for the two types of feedback. Explicit feedback will benefit participants more than implicit feedback.

6.3 Methodology

This experiment employed the pretest-treatment-posttest design standard in applied linguistics. During the pretest, subjects did a self-paced reading task to gauge their online processing strategies, completed traditional paper-and-pencil tests which included an interpretation test and a production test to measure their knowledge about anaphora. This was necessary to make sure that any gains on the posttests were not due to prior knowledge and to make sure that their knowledge about reflexives is less than perfect. For the treatment, participants learned about the grammatical constraints on English pronouns and reflexives through a computer program developed by the researcher. “Director”, software for developing computer-aided language learning, was used. The template for the training program was designed by Robert Ariew at the University of Arizona. The program recorded the responses and the length of the time that each subject spent on the tasks. Posttest assessment consisted of one interpretation test similar to the pretest, one sentence completion task (as the production test following most PI studies), and one self-paced reading task. The difference in scores between the pretests and posttests were subject to statistical analysis to determine whether the treatment was helpful. Participants’ background information was collected through a questionnaire. Participants’ perception of the
learning experience was collected via an exit survey. Specific instrumentation, participants and procedures for the experiment will be described in the following sections.

6.4 Instrumentation

In Experiment 3, five instruments were used. They are: self-paced reading tasks for the pretest and the posttest, offline written tests for the pretest and the posttest, a language profile, computer-delivered processing instruction as the treatment, and an exit survey of the treatment. The following sections describe the rationale and sample materials of each instrument.

6.4.1 Self-Paced Reading Tasks: Pretest and Posttest

Because Experiments 1 and 2 established that self-paced reading tasks were a valid tool to assess both native speakers and nonnative speakers’ implicit knowledge of anaphora, the same task was used in Experiment 3. As Experiment 3 was to test whether PI can help learners alter how they process anaphors, it was necessary that participants be assessed about their ability to employ the structural cues to assign antecedents. This is exactly what Experiment 2 aimed to do. Naturally, the same materials used in Experiment 2 were used for Experiment 3. Because Experiment 3 uses a pretest-posttest design, ideally, there should be two parallel self-paced reading tasks, one for the pretest and one for the posttest. However, there are too many factors
involved in the self-paced reading tasks such as the grammatical structure, vocabulary including word frequency, and word categories, that influence the reading patterns. This makes it almost impossible to construct two parallel tests. Therefore, the same materials were used for the pretest and posttest. One potential problem of this practice is item familiarity which means that participants may remember their responses to the items in the previous tests. One common strategy is to lengthen the time between the two events so that the effect is minimized. This is the approach in the current experiment. Pretests were conducted at least two weeks before the posttests.

Although results of Experiment 2 showed that the materials achieved the effect that was expected, some minor adjustment was made to accommodate the participants in Experiment 3. Several participants revealed after Experiment 2 that they were not sure whether some of the names were male or female. Since it is crucial that participants know the gender of the names, some names were changed to more familiar ones. For example, “Jill” was changed to “Mary”. All together, three names were changed. No other adjustment was made.

6.4.2 Offline Written Test: Pretest and Posttest

It is well known that it is extremely difficult to assess the knowledge of reflexives because the results may just reflect participants’ preference when there are alternative readings for a given sentence (e.g., White et. al., 1997). Different assessment tools such as multiple-choice tasks, i.e., asking subjects to choose from a list of potential
antecedents; picture-identification tasks, i.e., choosing from two or more pictures illustrating potential interpretations; and truth-value judgment tasks, i.e., giving stories as context for a sentence or giving pictures with sentences for subjects to judge whether they match or not, have been used in the literature. White, Bruhn-Garavito, Kawasaki, Pater and Prevost (1997) compared two truth-value judgment tasks, one involving stories and the other pictures, and found out that the former slightly better reflects L2 interlanguage grammar. One concern is that reading a large number of stories is demanding for low proficient learners. Another candidate is a multiple-choice comprehension test with biased sentences (used by Yuan, 1998; Thomas, 1989, 1993; Demirci, 2000). Those biased sentences provide context, similar to those stories, to the interpretation of the sentences. Sample stimuli are:

Subjects read: After the medical tests were completed, the doctor informed Bill about himself.

Subjects answer: Can “himself” refer to the doctor? Yes/no
Can “himself” refer to Bill? Yes/no
Can “himself” refer to somebody else? Yes/no

This is the method chosen for the current study. The interpretation test consisted of 12 sentences (See Appendix H). There were 9 critical sentences testing participants’ knowledge of anaphora. The other three were distractors. The design was essentially the same as the written test in Experiment 2: The materials in Experiment 2 were included in Experiment 3. However, in addition, a second version was created that
was balanced for difficulty and vocabulary. A split-block design was used such that one version was used for one group as a pretest and the other was used as a posttest. A complete list of the items is included in Appendix H.

Because previous PI studies also included production tests such as sentence completion tasks (e.g., Morgan-Short and Bowden, 2006), Experiment 3 also had a component of production test to make it comparable across studies. The production test included 4 items, with 2 critical items and 2 distractors (Appendix H). This test was a sentence completion task in which participants read a story and, following the story, completed a sentence with a pronoun which is a natural continuation or response to the story. Stories provided a context for the participants. Each story was 3 to 4 sentences long and easy for intermediate ESL learners to understand. Some stories were adapted from White et. al. (1997) by simplifying the stories and changing the task type from interpretation to production. The rest were created by the researcher. Participants have to understand the story in order to answer the questions correctly. Distractors were necessary so that they could not just fill reflexives in the blanks. The following are the sample stories:

Sample critical stimuli:

Once a week, Mike used to visit an old woman. On Mike’s last visit he saw the old woman point a gun at her head and shoot. The old woman died instantly.

Mike knew that the old woman shot______.

Sample distractor:

Killer Harry was free again. Bill was very scared. Bill called a policeman so the
policeman could guard him and make sure he was safe from Killer Harry.

Bill asked the policeman to protect ________.

6.4.3 Language Profile

The same language profile was used as in Experiment 1. See Appendix C for the questionnaire. It is necessary to gather participants’ language learning experience to make sure that the study group is more or less homogenous.

6.4.4 Computer-Delivered Processing Instruction: Treatment

The treatment was PI. As discussed in Chapter 2, normally PI includes three components: an explicit explanation; referential activities; and affective activities. The last two are usually called “structured activities”. To make sure that the PI activities designed in this study conform to the guidelines of PI as described in VanPatten (2004), the following will illustrate how the materials were created according to the six guidelines and principles.

The first guideline is: One thing at a time. Here one thing means one function or structure to be taught. In the activities, only “himself” and “herself” will be taught in the context of one important rule: English reflexives are locally bound.

The second guideline is: Keep meaning in focus. In each activity, understanding the sentence meaning is necessary for learners to successfully complete the tasks. They cannot use other cues such as gender information to find out the answer because
gender is deliberately made the same for the sentences in the materials. For example, in “Mr. Green said that Mr. Blue cut himself”, learners cannot single out the antecedent based on gender because both potential antecedents are male. If learners read “Mrs. Green said that Mr. Blue cut himself”, based on the gender information they could answer the question “who is himself?” without trying to understand the meaning of the sentence.

The third guideline is: Learners must do something with the input. In the PI training materials, learners were required to not only read or listen, but also to make a decision about what they had just read or heard. They had to either select a picture or select one from multiple choices that match or answer the question. For the referential activities, the answer was form-based and for the affective activities, the response was opinion-based and required them to express their beliefs and opinions.

The fourth guideline is: Input should be both oral and written. The training materials contained 2 oral activities in which learners listened to recordings and answered some questions by either selecting pictures or doing regular multiple choices. The training also contained activities that required learners to read and then answer questions. The treatment did not require learners to write anything because it is hard to code learners’ errors when they type on the computer. Errors could be typos or reflect a lack of knowledge. Instead, a sentence completion task was to simulate a written task. In this task, learners read a 3- to 5-sentence story and then completed a sentence according to the story. Instead of asking them to generate a response, they
were given choices.

The fifth guideline is: Move from sentences to discourse. The packet combined input activities both at the sentence level and the text level. In addition to some short stories, learners also had to read a short passage and make a decision about the target structure. This activity was at the end of the training.

The last guideline is: Keep the learners’ processing strategies in mind. This is probably the most important guideline since PI aims to alter learners’ incorrect processing strategies. Processing strategies is a general term which refers to what learners do when they attempt to understand a sentence during their reading. One such strategy is the First Noun Strategy, i.e., treating the first noun phrase as the agent of the action (Slobin and Bever, 1982; VanPatten, 1984). As found out in Experiment 2 in the written test, less proficient Chinese ESL learners allowed long-distance binding and also resorted to other cues such as gender and pragmatic meaning when locating antecedents for reflexives. Learners relied on semantic information rather than structural information. The pretest results of the online self-paced reading task in Experiment 3 (See Table 15) confirmed this observation, i.e., during online reading, learners allowed long-distance antecedents and underused structural information. In order to help learners overcome the strategy of over-relying on semantic information, in the activities here, other cues such as gender are not very helpful to single out the antecedents, thus forcing learners to analyze the sentence structure to find the antecedent.
As pointed out by Lee and VanPatten (1995), what learners are required to do can vary from one activity to another. VanPatten (2004) suggested possible activities such as multiple choice activities, matching activities, and others. The current packet presented 41 instances of the target form in both referential and affective activities such as multiple choice tasks, listening and story reading. The drawings that were paired with the sentences were either modeled after other studies (e.g., White et. al., 1997). The following are some sample items from one referential activity and from one affective activity. Appendix I presents the complete activities in the training packet.

Sample referential activity:

Instruction: Listen and then select the picture that goes with the sentence.

Voice stimulus: 1. Mr. Green said angrily that Mr. Blue hit himself.

Visual stimuli:
Sample affective activity:

Instruction: Select a female classmate or relative or friend of yours (mother, sister, aunt, niece, etc.) and keep her name in mind. Which of the following statements is likely true to her?

1. She says that her father will buy himself a new car.  True_____  Not True_____

As the experiment involved computer-delivered instruction, all activities were digitized. “Director” software was used to present the instruction in which learners followed the lesson by clicking the mouse. The researcher put all the activities in the program. In the final training program, learners first logged in using their pseudo names and pseudo-IDs as illustrated in the following screen shot, Figure 3.
Then learners clicked the arrow on the lower right screen to continue to the actual training. There were eight activities including matching pictures with recordings, reading sentences and then indicating who is performing the action, true or false affective activities, sentence completion, listening and then indicating the doer of the action, reading stories and then completing sentences, reading passages and answering questions. The following sample screen shots illustrate what the program is like.

First learners saw a screen as shown in Figure 4, and they were given instructions to click on the speaker icon to listen to the recordings. After learners clicked on the speaker, the program played the voice stimulus and then, as soon as it ended, two pictures were displayed on the screen along with instructions asking learners to choose one picture to match the recording. This is shown in Figure 5. When learners
clicked on one picture, the program provided immediate feedback as shown in Figure 6. The type of feedback learners got depended on what treatment group they were in, i.e., the explicit feedback group received an explicit explanation and the implicit feedback group was simply told “right” or “wrong” (See Figure 8 for the illustration). Time spent on each event was recorded from the moment when learners clicked on the speaker icon to the moment when they clicked on the arrow on the bottom of the screen. When one part was completed, a screen shot such as shown in Figure 7 appeared, and clicking on the arrow took learners to the next part.

Figure 4: Screen shot: Instructions for activity one
Do you know who is performing the action?

Click on the speaker icon to hear a recording. You may listen to it more than once by clicking on the icon.

Now click on the image that best represents the situation described in the recording.

Figure 5: Screen shot: After audio stimulus for activity one

Do you know who is performing the action?

Wrong

Figure 6: Screen shot: Feedback
There were two treatment groups. One group received implicit feedback with structured activities and the other group received explicit feedback with structured activities. There was no explicit explanation before the structured activities for either group. Implicit feedback was just yes/no feedback, informing participants whether their response was correct or not. Explicit feedback provided a metalinguistic explanation, i.e., yes/no and why the answer was wrong. Figure 8 shows a sample screen shot for the explicit feedback condition. This is a referential activity in which learners first listen to a recording and then select a picture that matches the recording. The recording sentence was: “Mr. Green said angrily that Mr. Blue hit himself.” This sentence was used to train learners that even if the sentence meaning favors the interpretation that “himself” refers to Mr. Green, the correct response is that “himself”
can only refer to the local noun phrase. The explicit feedback for a wrong response would be: “Wrong. Here ‘himself’ can only refer to Mr. Blue. The sentence meaning may make you choose the other picture, but grammar does not allow that interpretation. Keep in mind that English reflexives such as ‘himself’ and ‘herself’ are strictly locally bound, i.e., they refer to the person within the same phrase.”
6.4.5 Exit Survey of the Treatment

To find out how learners perceived the learning experience with a computer program, exit surveys were incorporated in the “Director” program. There were five multiple choice questions probing learners’ learning experiences with the interactive computer-delivered lesson. The survey provided a window for the researcher to understand what learners thought of the lesson and how enjoyable the lesson was and their opinions on the best way to learn anaphora. The survey questions are located in Appendix J. The following figures are sample screen shots for the survey.

Figure 9: Screen shot: Beginning of the survey
6.5 Participants

There were 30 adult ESL participants in Experiment 3. Two were excluded because they got perfect scores in the pretest. Each participant was compensated a small amount of money for his or her time. All participants’ first language was Chinese. Participants in Experiment 3 were not as proficient as in Experiments 1 and 2. This was necessary to make sure that they had not fully acquired the structure that was going to be taught. According to the language profile, they were intermediate ESL learners. Their average self-ratings of English proficiency in the four skills (speaking, listening, reading and writing) were 2.7, 2.7, 3.3, 2.8, respectively on a scale of 1 to 5 with 5 the best. Their age ranged from 19 to 40, with a mean of 28.
Their average TOEFL score was 589, ranging from 510 to 630; however, out of 28, only 11 reported their TOEFL scores. Participants started to learn English at an average age of 12, ranging from 7 to 15. All participants were well educated, with a mean number of years of formal education of 18 years, ranging from 12 to 23. Based on the education system in mainland China, this means that on average they had learned English at school for at least 10 years. Participants’ mean number of months of residence in the U.S. was 27, ranging from 1 to 72. Their mean years of education in English-speaking countries was 1.5, ranging from 0 to 6. The following table summarizes participants’ background statistics.

Table 13: Participants’ background in Experiment 3

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Minimum</th>
<th>Maximum</th>
<th>SD</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>28</td>
<td>19</td>
<td>40</td>
<td>5.2</td>
<td>28</td>
</tr>
<tr>
<td>TOEFL scores</td>
<td>589</td>
<td>510</td>
<td>630</td>
<td>39.0</td>
<td>11</td>
</tr>
<tr>
<td>Age starting English</td>
<td>11.9</td>
<td>7</td>
<td>15</td>
<td>1.6</td>
<td>28</td>
</tr>
<tr>
<td>Years of formal education</td>
<td>17.8</td>
<td>12</td>
<td>23</td>
<td>3.0</td>
<td>28</td>
</tr>
<tr>
<td>Months of residence in the U.S.</td>
<td>26.6</td>
<td>1</td>
<td>72</td>
<td>21.3</td>
<td>28</td>
</tr>
<tr>
<td>Years of education in English-speaking countries</td>
<td>1.5</td>
<td>0</td>
<td>6</td>
<td>2.0</td>
<td>28</td>
</tr>
<tr>
<td>Self-rating of English proficiency</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Speaking</td>
<td>2.7</td>
<td>1</td>
<td>5</td>
<td>0.9</td>
<td>28</td>
</tr>
<tr>
<td>Listening</td>
<td>2.7</td>
<td>1</td>
<td>5</td>
<td>0.8</td>
<td>28</td>
</tr>
<tr>
<td>Reading</td>
<td>3.3</td>
<td>2</td>
<td>5</td>
<td>0.8</td>
<td>28</td>
</tr>
<tr>
<td>Writing</td>
<td>2.8</td>
<td>2</td>
<td>4</td>
<td>0.8</td>
<td>28</td>
</tr>
</tbody>
</table>

6.6 Procedures

Participants were tested individually in the same setting as in Experiments 1 and
2. There were 3 steps. In step 1, participants first did the self-paced reading task as a pretest. This took them less than 35 minutes. After a 5- to 10-minute break, they completed the paper-and-pencil test which reflects their explicit knowledge of anaphora. The test took less than 20 minutes. Participants also filled in a language learning background questionnaire during the pretest (Appendix C). The questionnaire took less than 5 minutes.

In step 2, participants were trained through the interactive computer program as described in section 6.4.4. This procedure was conducted at least one week after the pretests. This interval was necessary to minimize the potential sensitizing effect of the pretests. Participants were randomly assigned to two treatment groups based on the order in which they arrived for the experiment. Time on tasks was recorded. The training took less than 35 minutes for all participants.

Step 3 was the posttests. Posttests were completed within one week after the treatment. Posttests were not administered immediately after the treatment because more time was needed between the pretest self-paced reading task and the posttest reading task to minimize the potential item familiarity effect. Also, it was important to determine whether the training had “lasting effects”. There were two parts to the posttests. First, participants did the self-paced reading task on the computer. This took them less than 35 minutes. Then they took a short break and finished a written test which was a different version of the one they had for the pretest, i.e., if a participant took version A as a pretest, they got version B as a posttest or vice versa. The written
test had no time limit but all participants took less than 20 minutes.

6.7 Data Analysis and Results

Before examining participants’ online RT data, their scores on the written pretest were calculated. There were two parts to the written test, i.e., an interpretation task and a production task. As described in Section 6.4.2, the interpretation task required participants to interpret anaphora by answering three yes/no questions for each item. There were 12 items with a possible 12 points. One point was awarded if participants answered the three yes/no questions correctly for each sentence. No points were awarded if any mistake was made. For example, for a sentence such as “Jack says that Peter talked about himself”, the correct answer is that “himself” can only refer to Peter. If participants indicated that “himself” here can refer to either Jack or Peter or only refer to Jack, they got a zero for the item. Raw scores were then converted to percentages by dividing by 12. The same scoring procedure was applied to the second half of the pretest except that production items had one question for one item. There were four items. So the possible number of points was 4. Those who scored below 95% at the interpretation test were included in the experiment. Two participants were excluded for this reason. The same scoring procedure was applied to the posttest. The following figures present the results of the pretest and the posttest for each individual.
Scores for the interpretation task: Pretest and Posttest

Figure 11: Scores of the interpretation task in the pretest and posttest for each individual in the implicit feedback group (The mean for pretest M=0.59; and the mean for posttest M=0.93)

Scores for the interpretation task: Pretest and posttest

Figure 12: Scores of the interpretation task in the pretest and posttest for each individual in the explicit feedback group (The mean for pretest M=0.56; and the mean for posttest M=0.90)
Scores for the production task: Pretest and Posttest

Figure 13: Scores of the production task in the pretest and posttest for each individual in the implicit feedback group (The mean for pretest M=0.98; and the mean for posttest M=0.95)

Scores for the production task: Pretest and posttest

Figure 14: Scores of the production task in the pretest and posttest for each individual in the explicit feedback group (The mean for pretest M=0.93; and the mean for posttest M=0.93)

The data were subjected to repeated measures ANOVA tests in SPSS. Tests of
within-subjects contrasts showed that the difference between pretest interpretation and posttest interpretation scores was significant, $F(1, 26)=64.1, p<.05$. The difference between pretest production scores and posttest production scores was not significant, $F(1, 26)=1.0, p=.32$. Tests of between-subjects effects showed that the difference between the implicit feedback group and the explicit feedback group in improvement in interpretation was not significant, $F(1, 26)=.39, p=.54$. The same was true of the production score improvement, $F(1, 26)=2.0, p=.17$.

Time on task is a potential variable for the training effect, i.e., the longer participants spend on the task, the better improvement they obtain. The “Director” program was designed to record the time participants spent on each item. For some unknown reasons, data for some participants for some items were not recorded in the output file. Instead of using the total time on the training task to compare the two treatment groups, it was decided to calculate average time spent on each item for each participant. The average time for the implicit group and the explicit group (N=14) was 23 seconds (SD=8) and 21 seconds (SD=6), respectively. Independent Samples T-test showed that the two groups did not differ in the average time they spent on each item, $t=0.6, df=26, p=0.53$.

The average comprehension error rate in the pretest was 26%, ranging from 15% to 40% (SD=0.07). The mean error rate in the posttest was 21%, ranging from 8% to 40% (SD=0.09). This suggested that, overall, participants understood most of the sentences. Data for the self-paced reading tasks were also analyzed using the SPSS
program. Participants’ RTs were trimmed before the actual statistical analysis. The same procedure was applied as in Experiments 1 and 2. Those RTs that were two SDs longer or shorter than the same participant’s mean, or higher or lower than the high and low cutoffs set at 2,000 and 200 ms, respectively, were excluded. These procedures, along with missing data and display errors, accounted for 10% of the data. Thirty-two mean RTs were computed for each participant, one for each test position (four test positions: positions 1 to 4) in each condition (four conditions: both match, only NP2 match, only NP1 match and no match) for each sentence category (the neutral sentences and the pragmatically biased sentences). The prediction was that any anomaly (in structure or meaning) would produce a slowdown. Therefore, the statistical test was deemed one-tailed. For the item analysis, as in Experiment 2, sixteen means were calculated, one for each of the four test positions in each of the four conditions. The participants’ RTs for each test position, each condition, and each sentence category are presented in the following. Data for the pretest and the posttest are presented below.
Table 14: Participants’ mean RTs (ms) at four test positions for neutral sentences and pragmatically biased sentences involving the structural cue for conditions “both match” and “no match” in the pretest

<table>
<thead>
<tr>
<th>Neutral</th>
<th>Biased toward NP1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test position</td>
<td>1 introduced</td>
</tr>
<tr>
<td>Both match</td>
<td>647</td>
</tr>
<tr>
<td>No match</td>
<td>656</td>
</tr>
<tr>
<td>Difference</td>
<td>-9</td>
</tr>
</tbody>
</table>

*significant at .05 in subject analysis.
**significant at .05 in both subject and item analyses.

Table 15: Participants’ mean RTs (ms) at four test positions for neutral sentences and pragmatically biased sentences involving the structural cue for conditions “only NP1 match” and “only NP2 match” in the pretest

<table>
<thead>
<tr>
<th>Neutral</th>
<th>Biased toward NP1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test position</td>
<td>1 introduced</td>
</tr>
<tr>
<td>Only NP1 match</td>
<td>666</td>
</tr>
<tr>
<td>Only NP2 match</td>
<td>665</td>
</tr>
<tr>
<td>Difference</td>
<td>1</td>
</tr>
</tbody>
</table>

**significant at .05 in both subject and item analyses.
Table 16: Participants’ mean RTs (ms) at four test positions for neutral sentences and pragmatically biased sentences involving the structural cue for conditions “both match” and “no match” in the posttest

<table>
<thead>
<tr>
<th>Test position</th>
<th>Neutral</th>
<th>Biased toward NP1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 introduced</td>
<td>2 self</td>
</tr>
<tr>
<td>Both match</td>
<td>559</td>
<td>601</td>
</tr>
<tr>
<td>No match</td>
<td>556</td>
<td>614</td>
</tr>
<tr>
<td>Difference</td>
<td>3</td>
<td>-12</td>
</tr>
</tbody>
</table>

**significant at .05 in both subject and item analyses.

Table 17: Participants’ mean RTs (ms) at four test positions for neutral sentences and pragmatically biased sentences involving the structural cue for conditions “only NP1 match” and “only NP2 match” in the posttest

<table>
<thead>
<tr>
<th>Test position</th>
<th>Neutral</th>
<th>Biased toward NP1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 introduced</td>
<td>2 self</td>
</tr>
<tr>
<td>Only NP1 match</td>
<td>580</td>
<td>607</td>
</tr>
<tr>
<td>Only NP2 match</td>
<td>555</td>
<td>622</td>
</tr>
<tr>
<td>Difference</td>
<td>25</td>
<td>-15</td>
</tr>
</tbody>
</table>

*significant at .05 in subject analysis.
**significant at .05 in both subject and item analyses.

The same statistical procedures were applied as in Experiment 2. Paired-Samples T-Tests were used to compare participants’ mean RTs for four conditions at four different positions and for two sentence categories. For each group, sixteen pairs were compared for subject analysis and item analysis, respectively. The details of the data
As can be seen from the above tables, in the pretest stage, the participants’ RTs at the first two positions showed no reliable differences for NNSs for both neutral and pragmatically biased sentences with the exception of the first position for the biased sentences. However, the difference was only significant by subject analysis but not significant by item analysis. This was expected because participants read essentially the same words in the sentences until they reached the word in the second position. In position 2, according to previous studies, RTs were expected to be the same due to the spill-over effect in self-paced reading tasks as discussed in Experiments 1 and 2. The results at the first two positions again confirmed the no-difference predictions.

RTs at the third and fourth positions showed different patterns for different conditions tested. RTs between the “both match” and “no match” sentences for position 3 for both the neutral sentences and the biased sentences were statistically significant both in subject analysis, \( t_1(27) = -2.0, p < .05 \) for the neutral sentences, \( t_1(27) = -2.2, p < .05 \) for the biases sentences, and in item analysis, \( t_2(47) = -2.0, p < .05 \) for the neutral sentences and \( t_2(47) = -1.6, p < .05 \) for the biases sentences. For the fourth position, the difference was not significant either in subject analysis, \( t_1(27) = -1.6, p = .06 \) for the neutral sentences, \( t_1(27) = .13, p = .45 \) for the biases sentences or in item analysis, \( t_2(47) = -1.1, p = .14 \) for the neutral sentences and \( t_2(47) = -.15, p = .44 \) for the biases sentences.
The comparison between “only NP1” and “only NP2” sentences yielded no significant results for both neutral and biased sentences in both positions. For neutral sentences, in position 3, the 28 ms difference was found nonsignificant for subject analysis, $t_1(27)=1.6, p=.06$ and for item analysis, $t_2(47)=1.2, p=.13$. The same was true for the fourth position, in subject analysis, $t_1(27)=-.44, p=.33$, and in item analysis, $t_2(47)=-.24, p=.41$. And for biased sentences, it was found that in position 3, the 13 ms difference was not significant in subject analysis, $t_1(27)=-.78, p=.22$, and in item analysis, $t_2(47)=-.65, p=.26$. In position 4, it was also found that the 11-ms difference was nonsignificant in both subject analysis, $t_1(27)=-.22, p=.42$, and in item analysis, $t_2(47)=-.14, p=.45$.

The results for the posttest stage showed a different pattern. As expected, RTs at the first two positions for both neutral and pragmatically biased sentences showed no reliable differences. RTs at the third and fourth positions showed different patterns for the planned comparisons. As in the pretest stage, RTs between the “both match” and “no match” sentences for position 3 for both the neutral sentences and the biased sentences were statistically significant both in subject analysis, $t_1(27)=-2.5, p<.05$ for the neutral sentences, $t_1(27)=-3.4, p<.05$ for the biases sentences, and in item analysis, $t_2(47)=-1.9, p<.05$ for the neutral sentences, and $t_2(47)=-3.0, p<.05$ for the biases sentences. For the fourth position, the difference for the neutral sentences was not significant either in subject analysis, $t_1(27)=-.39, p=.35$, or in item analysis, $t_2(47)=-.16, p=.44$. However, the difference for the biased sentences was found
significant in both subject analysis, $t(27) = -3.4, p < .05$, and in item analysis, $t(47) = -3.3, p < .05$.

Unlike the pretest, the comparison between “only NP1” and “only NP2” sentences in the posttest yielded a different pattern. No reliable differences were found for positions 1 and 2 for both types of sentences. For neutral sentences, in position 3, the 35 ms difference was found significant for both subject analysis, $t(27) = 2.4, p < .05$, and for item analysis, $t(47) = 2.0, p < .05$. In the fourth position, the 21 ms difference approached significance in subject analysis, $t(27) = 1.6, p = .07$, and was nonsignificant in item analysis, $t(47) = 1.3, p = .10$. And for biased sentences, it was found that in position 3, the 10 ms difference was not significant in subject analysis, $t(27) = .55, p = .29$, and in item analysis, $t(47) = .32, p = .38$. In position 4, it was found that the 31 ms difference was significant in subject analysis, $t(27) = 1.9, p < .05$ and approached significance in item analysis, $t(47) = 1.5, p = .07$.

To summarize, as expected, it was found that there were no RT differences in positions 1 and 2 for both the pretest and the posttest stages. In the pretest stage, the only significant RT differences were in positions 3 for both neutral and biased sentences in the comparison between the “both match” and “no match” conditions. No other RT differences were found significant. In the posttest stage, besides the significant RT differences in position 3 for both types of sentences in those conditions, the RT difference in position 3 for the comparison between the “only NP1 match” and the “only NP2 match” conditions for the neutral sentences was significant. And the
RT difference in position 4 for the comparison between the “only NP1 match” and the “only NP2 match” conditions for the biased sentences was significant in subject analysis and approached significance in item analysis.

As participants’ perception is important for language learning because it may influence their motivation and then change their language learning behavior, Experiment 3 also asked participants to complete a short survey about their perception of the learning activities. Survey data were not all recorded due to a bug in the computer program. The program failed to record participants’ choices in the output file. Learners were then asked to write their responses on a piece of paper. Sixteen learners’ data were included in the analysis. It was found that most learners (75%) thought the interactive computer-based lesson was more enjoyable than paper-pencil classroom activities. Nineteen percent reported “about as enjoyable” and six percent indicated “somewhat less enjoyable”. Eighty-eight percent of the participants strongly agreed or agreed with the statement that “Because this activity is presented in an interactive computer-based format, I was better able to understand the ideas and concepts taught in it”. Only twelve percent disagreed with the statement. Also, eighty-eight percent indicated that they strongly disagreed or disagreed with the statement: “Because this activity is presented in an interactive computer-based format, I was at a disadvantage because I do not have adequate computer skills”. Only twelve percent strongly agreed. When rating the usefulness of the explicit explanation of grammar rules in the lesson, eighty-eight percent of the participants chose “very
useful” or “useful”. When asked to indicate whether they agreed with the statement: “In order to learn reflexive pronouns better, I would like to have chances to produce or speak reflexive pronouns during the lesson”, eighty-eight percent strongly agreed or agreed. Only twelve percent disagreed with the statement. The survey suggested that participants enjoyed the lesson.

6.8 Discussion and Conclusion

Experiment 3 aimed to answer the following research questions: Can we alter L2 learners’ processing strategies with computer-delivered processing instruction? What is the role of explicit feedback in CALL in the acquisition of anaphora?

Experiment 3 found that in the pretest participants were not sensitive to the structural information when assigning antecedents. They did slow down when there was no match at all in the sentences for the antecedent as shown in the comparison between the “both match” and “no match” conditions. And this was true for both neutral sentences and biased sentences. This suggested that they have acquired some properties of anaphora. For example, this could mean that they used the gender information to single out one antecedent and when this failed, they encountered reading difficulties. However, they have not acquired the full properties of anaphora, i.e., the pattern of the reading time differences between the “only NP1 match” and the “only NP2 match” conditions showed that they were not sensitive to the structural information. Sentences in the former condition are ungrammatical and sentences in
the latter condition are grammatical. Based on the results of Experiment 2 and previous online studies, participants should spend more time reading the ungrammatical version. This is not the case. Participants in Experiment 3 showed no RT differences in all positions tested, which suggested that they treated both types of sentences equally. Their processing strategy was different from the NSs and the advanced ESL learners as in Experiment 2. Those sentences in which the only noun phrase that matched the gender of the anaphora was the inaccessible one were treated as acceptable. This could be attributed to the fact that in the L1, such sentences are legitimate. More analysis on this aspect will be discussed in the Discussion and Conclusions chapter of this dissertation.

After the computer-delivered processing instruction, results showed that participants improved their performances significantly as measured by the offline written tests. Most of them disallowed long-distance antecedents and allowed only local binding even under the pressure of pragmatic content favoring the alternatives. Although the production tests did not show evidence of improvement, it may be due to the ceiling effect. Participants had already scored high in the pretest, with an average of 96%. For the interpretation tests, participants improved significantly. This is a significant finding since previous intervention studies (e.g., White, 1995) failed to help learners acquire the feature. Most importantly, the online findings showed that ESL learners, to a certain extent, altered their processing strategies after the treatment. In the posttest, the RT differences were significantly different between the “only NP1
match” condition and the “only NP2 match” condition, which is a critical improvement in comparison to the pretest results. This means that participants have changed their language behavior from treating the two conditions as the same to treating them differently. Although there were differences between neutral sentences and biased sentences, i.e., results were more robust for the neutral sentences and less so for the biased sentences. For the biased sentences, the effect appeared rather late, i.e., only in position 4, and was significant in subject analysis but only approached significant in item analysis. This suggests that biased sentences were harder to process. This is understandable because participants have to overcome the pressure of the semantic meaning of the sentences which favors the wrong interpretation, and meanwhile they need to resist the temptation to allow long-distance binding. This finding is significant because it indicates that PI is effective not only as measured by traditional offline tests, but also effective as measured by online psycholinguistic methods. This topic will be revisited in the last chapter.

The role of feedback in CALL was examined in Experiment 3. Results showed no significant difference in improvement between the group who received explicit feedback and the group who received implicit feedback. Both groups improved in the interpretation task and also improved in the online self-paced reading task. This failed to confirm the hypothesis that the explicit group would benefit more than the implicit group. Possible reasons will be discussed in Section 7.2.3, Chapter 7.

To summarize, Experiment 3 was designed to determine whether PI could alter
L2 learners’ processing strategies and examine the role of feedback in CALL in the acquisition of anaphora. Findings showed that computer-delivered PI did alter L2 learners’ processing strategies and also it improved participants’ performances in the offline tasks as demonstrated in numerous PI studies. Results also indicated that explicit feedback in CALL was not necessary in the acquisition of anaphora if the instruction was PI.
7.1 Introduction

The purpose of this dissertation was two-fold. The central question that the research investigated was what information or strategies L2 learners use to comprehend pronouns and reflexives as sentences unfold over time. Meanwhile, the study looked into whether Processing Instruction was effective for the development of online processing strategies. The role of feedback in CALL was also investigated. Three separate experiments were carried out to investigate the above issues. Below is a summary of the findings and their implications for issues in second language acquisition such as the development of processing routines, transfer, knowledge representation and implications for language pedagogy such as the effectiveness of Processing Instruction and the role of feedback in language teaching. Limitations of the study and suggestions for future research will also be discussed.

7.2 Synopsis of Main Findings and Discussion

The following sections will provide a synopsis of the major findings experiment by experiment. Issues pertaining to the findings will be raised and discussed in turn.
7.2.1 Experiment 1

In Experiment 1, it was found that English learners, compared to native speakers, were sensitive to some but not all information of the pronouns and reflexives when reading sentences for general comprehension. Advanced L2 learners used the gender information automatically, i.e., any mismatched gender of the pronoun and its antecedent disrupted their reading process. However, unlike native speakers of English, L2 learners ignored the number information, i.e., when there is a mismatch in number between a reflexive and its antecedent in the sentences, their reading processes were not interrupted. It was also found that L2 learners were highly sensitive to verb information. When there is a mismatch between an NP1-biased verb and its preferred pronoun, such as in “The mother amused the father because he told funny jokes”, L2 learners’ reading was disrupted immediately following the unexpected word “he” because here “amused” is a verb which prefers the first NP as the coreference. The same was true to those NP2-biased verbs such as “envy”.

As argued in Section 4.8, Chapter 4, those findings are significant because all participants performed perfectly when tested offline using a traditional paper-pencil test. The discrepancy in performance between the online test and offline test points to the advantage of using different methods in language testing and research. Had the study used only one method, the conclusion would be very different. If only using the offline test, the study would have concluded that learners had acquired the target structure. And if only the online method was used, the study would have
underestimated the nature of the learners’ interlanguage. As discussed in Section 3.3.1.2, Chapter 3, by using both methods, the current study avoided the problems associated with them. Following Jiang (2004), the online method is taken as a tool to detect automatic processing competence by L2 learners, i.e., whether they have integrated the target structure in their mental representation. If learners can automatically use the target structure during online tasks, it is assumed that they have acquired the automatic processing routines of the target structure. Learners may know the rules as shown by offline tests but they may not have developed the appropriate processing routines as shown by the online tests. Automaticity in processing is very important in learners’ fluent use of the language.

As discussed in Section 4.8, Chapter 4, the findings of Experiment 1 led to the conclusion that ESL learners’ automatic competence is selective. For the same target structure, learners are sensitive to some cues, such as the gender information, but not others such as the number information. The reason automatic competence is selective remains a mystery, however. Previous studies (Broeder, 1995; Felix and Hahn, 1985) found a staged development in pronouns. Form-meaning connections are established sequentially based on case, number, person and gender, based on the rank order of the error frequencies learners produced. The gender feature is the most difficult, yielding 48% of all the substitution errors. The number feature errors only accounted for 10% of the errors. Although the corresponding order of acquisition is not certain because the data consisted of only errors, it does suggest that the number feature is not the
most difficult one. Owing to a lack of comparable data about reflexives, it is still an open question whether or not this developmental sequence is also true for reflexives. The current study showed that the number feature is difficult to be acquired and the gender feature is acquired early in terms of processing routines.

Besides the possibility that the plural feature is special because it poses great difficulties to L2 learners as discussed in Section 4.8, there are several other possible accounts for why learners failed to acquire the number feature. They are: the transfer account, the saliency account, and the frequency account. But none of them offer a completely satisfactory explanation.

It is generally accepted that L1 shapes the learning of the L2 grammar (White, 1985, Vainikka and Young-Scholten, 1996; Hawkins and Chan, 1997). Transfer has been extensively investigated both from a formal linguistics point of view and from a psycholinguistics perspective. The formal approach focuses on parameters setting and resetting under the framework of UG (e.g., White, 1985). The latter approach emphasizes processing strategies (e.g., Dussias, 2003) and low-level processing such as concurrences and speech segmentation. Will the transfer account, i.e., learners use the same information as in their L1, explain the mystery? It seems unlikely. Chinese, the L1 of the ESL participants, distinguishes the number feature of pronouns and reflexives by adding “men” to the singular form. For example, “ta” means the third person singular while “ta men” refers to the third person plural. If L2 learners applied their L1 processing strategy, they should be able to detect the anomaly in those
sentences automatically. However, a caveat is in order here. In Chinese, plurality is not marked in other structures as in English. For example, there is no plural marker for nouns. In fact, Jiang, Masuda, and Wang (2008) argued that the plural feature is not grammatized for Chinese speakers whereas it is for English speakers. According to them, “a concept/meaning is grammatized if there is a morphosyntactic device to express it” (Jiang, et. al., p.35). They proposed the Morphological Congruence Hypothesis, which claims that “a rule-governed L2 grammatical morpheme is acquirable only when the same meaning is grammatized and thus a similar morpheme is instantiated in the learners’ L1.” (Jiang, et. al., p.4). They further claimed that there is a critical period for the development of the grammatization of meaning. Assuming that the plural feature for participants in the current experiment is not grammatized in their mental representation, this may lead to difficulties in activating the actively represented meaning of plurality. There is some evidence that shows speakers of a language with number marking such as English, and speakers without such as Yucatec, behaved differently when judging the difference between the number of countable objects and the number or amount of non-countable matter. The former treated the differences in the number of objects more significantly, while the later group did not show such difference (Lucy, 1992). Jiang et. al. (2008), using a sentence-picture matching task, showed that semantic activation of the “plural” meaning was different for groups of different L1s such as English, Russian, Spanish, Japanese and Chinese. They found that English native speakers and Russian and
Spanish ESL speakers activated the “plural” meaning, but Chinese and Japanese ESL learners did not. In both Chinese and Japanese, plurality is not overly marked. This seems a plausible account; however, to make it work, it needs another assumption, which states that even though Chinese marks plurality in pronouns, somehow it is not a grammatized concept in their mental representation. The pronoun plurality marker of “men” in Chinese is only a surface form and it is only a lexicalized representation in the conceptual system. Or “men” is a special case to mark plurality of pronouns and reflexives and Chinese ESL learners do not link “men” with the English plural marker “-s”. Without further evidence, it is hard to accept this assumption. Besides, surface transfer, which includes low level transfer such as surface word order and morphologically similar gender marking transfer, usually is found easier than deep transfer, which is the transfer of more abstract features such as transfer of abstract syntactic categories (Sabourin, Stowe, and de Hann, 2006). Presumably, it should be easy to transfer from the similar surface form of the Chinese plural “men” to the English plural feature of pronouns and reflexives. This seems to go against the grammatization account.

Can perceptual salience explain the difficulty of the number feature? Perceptual salience has been proposed to influence the form-meaning connection between a form and a function when learners learn a language. The general assumption is that salient forms are easier to associate with functions than non-salient forms. For example, VanPatten (1996, 2003) explained why learners make certain connections, but not
others, by positing processing mechanisms. According to him, more meaning-bearing forms such as content words are first connected with their functions because they are essential for the message. Less meaning-bearing forms, such as the past tense marker –ed in English, may not draw learners’ attention initially because they are often redundant. For example, in “Yesterday I walked my dog”, the –ed does not add additional meaning and learners may skip it. VanPatten argued that, due to a limited processing capacity, such forms will be processed only when learners do not need attentional resources to process the content of the message. According to this account, lexical items are easier to establish form-meaning connections and grammatical forms are difficult. This can explain why verb information is automatically used by both L2 learners and native speakers. But it is hard to explain why there is a staged development of the gender feature and the number feature which are both morphosyntactic features. Without further assumptions, it is impossible to argue that the gender feature is more meaning-bearing than the number feature.

DeKeyser (2000) also posited that perceptual salience led to differential effects of age on the learning of certain structures. In his grammaticality judgment study, he found that both late L2 learners and early L2 learners obtained a high accuracy score on three structures, i.e., the basic word order, do-support in yes-no questions and pronoun gender, whereas both groups performed differently in other structures tested, such as the use of articles and the use and position of auxiliaries. He suggested that “pronoun gender errors are so irritating to native speakers that they will almost always...
correct them when their nonnative interlocutors make such mistakes, even though overt correction of grammar errors is otherwise rare in adult native-nonnative interaction” (DeKeyser, 2000, p.516). If this account is true, it can explain why learners in Experiment 1 acquire the gender feature of reflexives early. But it is difficult to argue that the number feature of reflexives is nonsalient. The difference between “himself” and “herself” is no more salient than the difference between “himself”/”herself” and “themselves”. This raises the issue of the difficulty of defining and operationalizing the concept of salience. DeKeyser (2000) used the chance of being corrected by native speakers as a criterion for salience for the pronoun gender errors, but used error positions for the other two structures on the assumption that sentence initial and sentence final positions are more salient. Salience is a complex construct that needs further research. For example, it needs to elucidate the roles of various factors such as the acoustic property, the context of a given message and the length of the target in perceptual salience. Goldschnieder and DeKeyser (2001) incorporated three components in the acoustic salience of inflectional morphemes: number of phones, syllabicity, and sonority. They posited that the more number of phones and the more sonorous in the morpheme, the more likely it will draw learners’ attention. This may be true for other morphemes and features such as the morphosyntactic features investigated in the current study. However, it is still unclear why the number feature is more difficult to process than the gender feature.
The last possible explanation is the frequency effect. Frequency has been proposed as an important factor in SLA. N. Ellis (2002) posited that frequency is a necessary and key factor in theories of language acquisition, and argued that language acquisition is exemplar-based and associative in nature because language rules from phonology to syntax “are structural regularities that emerge from learners’ lifetime analysis of the distributional characteristics of the language input” (N. Ellis, 2002, p.144). This model can account for the finding that advanced L2 learners can process the verb information automatically. When a given verb such as “amuse” is used in a context implying a causal relationship between two clauses linked by “because”, the subject, not the object of the verb, is a preferred coreference to start a clause with “because”. However, it encounters difficulty when it is used to explain the differential effects between the gender feature and the number feature. The input is consistent in the rules for reflexives. The gender and the number feature of the reflexives and antecedents are matched in the instances in input. Participants in Experiment 1 have extensive exposure to English and they should have extracted the patterns of those two features. It is possible that frequency does not affect all aspects of language in the same way. For instance, grammatical forms are known to reflect abstract underlying patterns and the rules may be too complex and require explicit learning. Besides, the frequency account is not without problems itself. Gass and Mackey (2002) pointed out some well-known phenomena in SLA. For some structures, language acquisition proceeds according to a certain sequence regardless of the input frequency such as the
acquisition of question formation (Lightbown and Spada, 1999) and the notoriously
difficult structures for L2 learners such as the article and the third-person singular –s.
Those phenomena do not necessarily rule out the role of frequency, but clearly they
raise issues that frequency may interact with other factors such as the nature of the
“rules” and developmental readiness in Pienemann’s (1998) term, i.e., acquisition
may occur only when learners have established the necessary processing capacity for
forms at the current stage of acquisition. For example, according to Pienemann’s
(1998) hierarchy of processing procedures, the long-distance binding principle
belongs to the “S-procedure” because it involves interphrasal information exchange.
The phrasal procedure, which involves only phrasal information exchange, is lower in
the hierarchy than the NP agreement.

7.2.2 Experiment 2
In Experiment 2, it was found that those L2 learners who have mastered the
English binding constraint of reflexives, as measured by a written test, also showed
sensitivity to the constraint during their online reading. When there was a mismatch
between a reflexive and its accessible antecedent in gender, learners took longer to
read the region immediately following the anomaly word. Furthermore, the
manipulation of pragmatics did not change learners’ reading patterns. For those
pragmatically biased sentences such as “The hungry girl was happy that the father
bought herself an ice-cream”, learners automatically detected the anomaly despite the
biased pragmatic information. Chinese, the participants’ L1, which allows both long-distance binding and local binding, did not influence their online reading pattern. The findings showed that those learners automatically used the structural information when comprehending sentences, and they did not use the L1 processing strategy for their L2 processing.

As discussed in Section 5.8, Chapter 5, findings in Experiment 2 were significant because they showed that learners can not only use semantic information such as the gender and verb information as demonstrated in Experiment 1, but also can automatically use structural information such as the binding principle for the moment-by-moment online sentence comprehension. This does not support the Shallow Syntax Hypothesis (Clahsen and Felser, 2006), which claims that L2 acquisition is fundamentally different from L1 acquisition and L2 processing is syntactically “shallow” and their syntactic representations computed during sentence comprehension is not detailed. L1 comprehenders may resort to shallow processing, i.e., without fully analyzing the deep structures of a sentence, and still understand a sentence. So shallow processing is not limited to L2 comprehenders, but L2 comprehenders are constrained by shallow structural computation. As discussed in Section 5.8, the current study showed that learners computed the deep structures of the sentences. Binding is established via the exchange of information between two phrases, which is not shallow. Comprehenders need to analyze the sentence structure to establish the coreference and cannot just rely on lexical information. The following
is a schematic representation of the online coreference processing mechanism that illustrates the complexity of the computation:

The above schematic sketch assumes that comprehenders activate a candidate set when assigning antecedents to “himself”. Both structural and featural information are used to eliminate inappropriate NPs. When comprehenders process a sentence, the
grammatical information of each word is accessed and stored. When encountering a reflexive or pronoun as in the example, comprehenders check the grammatical information associated with each NP and see if the features match those of pronouns and reflexives. Pragmatic information will come into play if there are more than one accessible NPs. In the above example, although both NPs, “son” and “father”, possess the same featural information, i.e., +male, +singular, as the reflexive “himself”, the structural information, i.e., the first NP “the son” is not within the phrase boundary of the reflexive and long-distance binding is not allowed in English, making “son” inappropriate as one of the candidates. If learners only rely on lexical information and do not use the structural information, they would treat “son” as a possible antecedent. The findings in Experiment 2 showed that this is not the case. Learners used the structural information automatically and so did the native speakers. Recall that in participants’ L1, Chinese, both NPs can be an appropriate antecedent. Pragmatic information can help readers resolve the ambiguity, but still both NPs are appropriate antecedents. In this case, readers will consider both of the NPs. The findings showed that learners did not apply the L1 strategy. Further, the strategy is unchanged if the more distant subject is the more plausible antecedent as in “The famous actress heard that the popular hostess talked about herself on TV”. Therefore, the current findings do not support the Shallow Syntax Hypothesis. Advanced English learners use both semantic and structural information to process sentences.
7.2.3 Experiment 3

Experiment 3 trained learners in the binding principle to see if they can acquire the online processing strategies after the specially designed structured activities. It also tested the role of feedback in CALL. Findings showed that learners improved their performances on the target structure significantly as measured by the offline written tests. Most of them disallowed long-distance antecedents and allowed only local binding. Furthermore, results showed that learners developed the appropriate processing strategies to a certain extent after the treatment. It showed that PI is an effective pedagogical technique that helps learners develop language-specific online processing routines and can lead to gains in the interlanguage developing system, not just gains in explicit knowledge. As for the role of feedback, results showed that explicit feedback was not necessary because gains for both the implicit group and the explicit group were not significantly different.

Experiment 3 adds to the growing literature attesting to the effectiveness of PI. It is one of the first to document that PI was not only effective as measured by traditional offline tests, but also effective as measured by online tasks. It confirmed previous studies’ claims that PI was effective (e.g., Marsden, 2006; Salaberry, 1997; Sanz and Morgan-Short, 2004; Takimoto, 2006; Toth, 2006; VanPatten and Sanz, 1995; VanPatten, 2002, 2005).

Previous studies such as White (1995) found that teaching the binding principle was not successful. In her study, as discussed in Section 2.3.3, the two treatment
groups did not improve at all after a four-week-long treatment with three 20-minute sessions per week. Although the failure may be due to the fact that they did not directly teach the rule, i.e., long-distance binding was not allowed, it is still puzzling given that learners had the chance to induce the rule because they were exposed to large numbers of exemplary sentences. Another study by White et. al. (1996) taught the long-distance binding directly to a group of Japanese learners. This time they were told that in Japanese the antecedent of “zibun” can be outside of the clause, but they were not taught that the antecedent must be a subject. Results showed that out of twelve participants, two French, one English, two Chinese and two Koreans knew the rule after the treatment. In Experiment 3, the implicit group only received yes/no feedback; they improved significantly, as shown by both the offline tests and the online tasks. This group is similar to the reading group in White (1995). The explicit group in Experiment 3 is similar to the other treatment groups in White (1995) and White et. al. (1996). Therefore, the success of the current experiment should be explained.

The difference in instruction may explain the different outcome. The current experiment used PI, a psycholinguistically based pedagogy, to teach the participants, while the other studies used traditional instruction. As discussed in the literature review, Section 2.5, many studies have shown that PI is effective for various target structures. Many studies attested that it was more effective than TI. The treatment in studies conducted by White (1995) and White et. al. (1996) could be classified as
traditional instruction given that they provided learners regular oral and written exercises ranging from mechanical to communicative. The tasks were not structured activities as in the current study. They did not push learners away from inappropriate processing strategies. Learners need to establish form-meaning connections when acquiring a target structure. PI helps learners form new connections and avoid old ones from their L1 that are not appropriate for the new language. To recap the key differences between PI and TI in Section 2.5.2, PI aims to help learners get intake from input using design activities that make form-meaning connections between nonsalient forms and function salient forms. PI activities also force learners to alter their inappropriate processing strategies for the target language. It emphasizes the important role of input in language acquisition while traditional instruction emphasizes the role of output-based exercises. TI activities do not force learners to change their processing strategies although they do sometimes inform learners about the difference between L1 and L2.

Experiment 3 found that feedback did not make a difference in learners’ gains in the binding principle as measured by both offline tests and online tasks. It confirmed the findings in many studies (e.g., Lyddon, 2007; Sanz, 2004; Sanz and Morgan-Short, 2004). For example, Sanz and Morgan-Short (2004) found that in their computer-assisted PI study, feedback did not matter. Lyddon’s (2007) computer assisted study also found no difference among different types of feedback groups. It did not support the claim that explicit feedback benefited learners more than implicit
feedback as made in many studies (e.g., Carroll and Swain, 1993; Carroll, 2001; Nagata, 1993; Rosa and Leow, 2004). Nagata (1993) proposed that metalinguistic feedback is more effective if a structure is complex, such as particles. The results do not support the claim. The binding principle is arguably complex since learners must analyze the phrase structures and it is classified as an inter-phrasal information exchange in Pienemann’s (1998) hierarchy of processing procedures. Given that all participants received the same structured activities, it is logical to conclude that structured activities were responsible for the gains. Explicit feedback did not enhance the learning. It is possible that learners induced the rule themselves or used the exposure-based learning mechanism in the current study.

### 7.3 Implications for SLA and Language Pedagogy

The findings in the current study have several major implications for SLA and language pedagogy. First, it is necessary to consider at least two factors in the development of language proficiency; one is quantitative and the other is qualitative. Traditional studies focused on the former and often used only offline written tests to gauge linguistic knowledge or explicit knowledge of learners. Psycholinguistic studies afford us a new tool to gauge not only the quantitative aspect, but also the qualitative aspect of proficiency. Automatic competence is part of the qualitative aspect of language proficiency. The current study showed that language learners’ development in automatic competence and processing routines is selective and they are not
equivalent to their linguistic competence. Using only offline measurement cannot capture the whole picture of learners’ language proficiency. Developing appropriate processing routines or automatic competence is important for learners in order to use or comprehend a second language as efficiently as L1 speakers do. According to Jiang (2007), there are three possible sources of knowledge integration: L1 knowledge via transfer; input and interaction via knowledge construction; and explicit knowledge via knowledge transformation. The outcome of knowledge is represented as integrated knowledge, similar to implicit knowledge. Automatic competence is the processing outcome. The current study, using a standard psycholinguistic method, demonstrated that it is important to examine the status of knowledge representation, i.e., whether it is implicit or explicit. Methodologically, the current study implies that psycholinguistic methods such as self-paced reading tasks are promising tools for the field of SLA. Future L2 studies must study both representation, processing and linguistic knowledge.

Second, language instruction is more beneficial if instructors focus on the processes that lead learners to turn input to intake, and push learners away from the wrong processing strategies. When instructors keep learners’ processing strategies in mind and design activities that are structured to help learners avoid the wrong strategies, language teaching is successful. In Experiment 3, learning activities were designed according to PI and those structured activities forced learners to focus on processing the input. This change of focus helped learners make new form-meaning
connections and made a difference in learners’ interlanguage development. According to VanPatten (2004), learners tend to use content words and other lexical items, rather than grammatical forms, to get meaning. Activities should be designed to encourage learners to process grammatical forms in order to correctly understand sentences. This will help learners acquire the correct processing strategies for the target language.

Third, language instruction should be input-focused instead of output-focused, especially at an early learning stage. PI was successful in helping learners acquire the binding principle in this study mainly because it focused on the input, not the output. It is not necessary to force learners to produce the target structure, especially at the initial stage of making form-meaning connections. This study supports the claim that input is important in language instruction. This is not to deny the important role of output in SLA. Teachers should not discard output entirely. Swain (1995) proposed three functions of output: the noticing/triggering function; the hypothesis-testing function; and the metalinguistic (reflective) function. Output helps learners notice the gap between what they say and what they should say; it helps learners test their assumptions about a particular language structure; and it also helps learners reflect what they know and discuss language phenomena. Furthermore, output promotes fluency, which is an important goal for any serious language learners. However, the role of input should be emphasized because many instructors now ask learners to perform communicative tasks without giving learners enough input or language exposure first.
Fourth, pedagogy must consider learners’ goals and provide ways to achieve those goals. If automaticity is an important goal to learners, instructors should help learners develop their automatic competence. PI is one way to do so because the outcome is not just explicit linguistic knowledge, but also implicit knowledge that can be deployed for automatic processing. It is well-known that often learners can articulate a grammatical rule but they cannot use it during their spontaneous speech. This is due to learners’ lack of implicit knowledge. Unfortunately, most current assessment does not test learners’ implicit knowledge or automatic competence. Teachers should use PI to cultivate learners’ implicit knowledge.

Fifth, instead of spending too much time designing explicit feedback in CALL, instructors should focus on designing useful learning activities. The current study implies that it is important to teach one thing at one time. Instead of teaching learners the whole conjugation of a target structure, it works best to focus on one point each time. It is also useful to keep meaning in focus so that learners can establish form-meaning connections while doing those activities. Other implications for designing appropriate activities are: Good activities should force learners to process structures that are not salient and those structures have little communicative value to get the meaning. Otherwise, learners will perform the exercises without developing the correct processing strategies for the target language. The current study suggests that feedback is not necessary; however, the findings do not imply that we should get rid of explicit feedback completely. Different learners may have different preferences.
or beliefs, which in turn impact their learning motivation and behavior. Explicit feedback has some benefits. Schulz (1996, 2001) found that learners liked and expected corrective feedback. Good and Brophy (2000) reported that most students found it psychologically rewarding to get immediate feedback. Lyddon (2007) also found that learners liked and expected feedback despite the fact that it did not make a difference in the learning outcome. We could give learners choices but our focus should be learning activities that lead to learners’ interlanguage development.

Last, but not least, learners are motivated to learn via computer-assisted language learning activities, especially when the program is interactive. After the treatment, many learners reported that the program was interesting and, especially with the audio, the pictures made the lessons memorable. When designing CALL activities, it is important to use multimedia to make them engaging and memorable. According to Mayer’s cognitive theory of multimedia learning (1996), learners possess two information processing systems: the verbal system and the visual system; and learners process a multimedia presentation of words, pictures and audio stimuli in an integrated fashion, i.e., each system builds a mental representation and interacts with each other. It is better to present words and pictures than words alone. When presented with both words and pictures, learners generate two representations using both the verbal system and the visual system and form connections between them. This leads to better information processing and learning.
7.4 Limitations of the Present Study

Several limitations of the current study merit discussion. For example, one limitation is that the language production test used was a “fill in the blank” type of test. Although previous PI studies used the same method and assumed that it was more or less similar to a production task, it is just an approximation. Unfortunately, a spontaneous language production task is hard to administer with a computer program because of the difficulty of automatic speech recognition. It cannot give any feedback if the speech cannot be recognized. For this reason, a sentence completion task was used. Therefore, the results provide little information about learners’ improvement in spontaneous language production.

A second drawback is that no long-term effect of the treatment was investigated. The results can only inform us about the short-term effect of PI. Unfortunately, it is very time-consuming to test long-term effects. In the current study, each participant in Experiment 3 had already spent more than 90 minutes on the task. To give another delayed test weeks later would be very difficult to administer. However, previous PI studies (e.g., VanPatten and Fernandez, 2003) showed that the effect can last up to eight months. It is not unreasonable to speculate that the effects could last for the current learners. Future studies are needed to measure the long-term effect of PI to confirm the speculation.

A third methodological limitation is that no comparison was made between PI and other types of instruction such as TI. No claims could be made about the relative
effectiveness among different types of instruction. Although the primary purpose of the study was not to compare different types of instruction methods, it would be interesting to see if other types of instruction, such as TI and meaning-focused instruction, could lead to the same gains in both offline tests and online tasks. Previous PI studies did show that PI was superior to TI. However, only offline tests were used. Future studies are necessary to compare PI and other types of instruction using the same offline and online tests in order to measure learning effects.

7.5 Suggestions for Future Research

Future research could compare PI and TI or other currently used teaching methods in mainstream classrooms such as meaning-focused instruction or task-based approach to see if PI is superior. If using the TI can also lead to the same gains in the target structure of pronouns and reflexives, caution should be exercised when advocating the use of PI over other instructional methods.

Various languages and target structures have been used to see if PI is effective. For example, PI has been used to teach Spanish pronouns (VanPatten and Cadierno, 1993), Italian future tense (Benati, 2001), and English simple past tense (Benati, 2005). This study studied English reflexives and attested PI’s effectiveness. It would be interesting to use participants with different L1s to see if the effect is language specific. For example, it is interesting to test participants with L1s whose binding
principles are similar with English, such as Spanish, to see if PI is still effective and to see if participants transfer their L1 strategies to their L2 processing.

The results point to the difficulty of processing the plural feature of reflexives and pronouns by Chinese ESL learners. Several possible explanations were discussed, but none is satisfactory. Future studies could pursue the grammaticalization account to find evidence that for Chinese speakers, the plural feature is not grammaticalized.

This study is one of the few to use psycholinguistic methods to study L2 processing. It showed that processing routines do not equal linguistic knowledge. It is important to document to what extent L2 learners could process sentences as L1 speakers do, and find out how to help L2 learners develop such processing abilities. This study revealed L2 learners’ processing strategies in comprehending pronouns and reflexives. There are many other structures that remain to be investigated such as the relative clause, and prepositional phrase attachment.

Finally, this study used self-paced reading tasks to gauge learners’ processing strategies. There are many other psycholinguistic methods such as eye-tracking and ERP. They are more time sensitive and could pinpoint the time course that L2 learners use various information to comprehend a sentence online. For example, it is still under debate when exactly the binding principle is applied during online sentence processing. Nicol and Swinney (1989) proposed the initial filter model, i.e., the binding principles were an initial filter for the possible antecedent candidates. Others disagreed. For example, Badecker and Straub (2002) proposed the
interactive-parallel-constraint model. For them, the binding theory, discourse focus and other constraints are simultaneously applied at the earliest stage during coreference processing. The initial candidates compose all prominent discourse entities that match the pronoun or anaphor in number and gender, analogous to accessing multiple meanings of an ambiguous lexical item. Using more time-sensitive methods, and using L2 learners as well as L1 participants may help us pinpoint the timeline and locate any differences.

7.6 Conclusion

This study used psycholinguistic methods to study online coreference processing. The results are encouraging for SLA. They showed that the moving-window self-paced reading technique is a valid tool for SLA, especially to investigate the status of knowledge representation, whether explicit or implicit. Results revealed that advanced L2 learners could use the same strategies as L1 speakers for some features, but not for others. Findings also showed that ESL learners used both semantic information and structural information to understand sentences as they unfold. It suggests that L2 processing is not “shallow”.

With regard to the intervention study, it was found that PI is an effective pedagogical technique that can alter learners’ inappropriate processing strategies, and lead to gains in interlanguage developing system, not just explicit knowledge. Finally, it was found that feedback is not necessary in CALL if activities are structured and
follow the guidelines of PI. Instructors could focus more on the quality of activities than on the types of feedback.

Although the study showed that PI is effective, it only tested one target structure in one language. Future studies will need to test more structures and more groups of learners with different L1s to establish the extent to which the findings can be generalized. Also, future studies need to address whether PI is superior to other instructional methods such as TI if both offline methods and online methods are used to assess the treatment effects.
APPENDIX A: ONLINE TEST STIMULI IN EXPERIMENT 1

Note: “+” indicates a positive answer and “−” indicates a negative answer.

1. The careless pedestrians found themselves/himself covered with mud.
   -Were the pedestrians covered with water?

2. The nervous actors calmed themselves/himself down after work.

3. The famous actresses prepared themselves/herself to face the crowd.

4. The dirty soldiers cleaned himself/themselves before going to bed.

5. The emotional sisters felt themselves/herself getting sentimental after the speech.
   -Did the sisters feel happy?

6. The shy girls forced themselves/herself to sing the part at the concert.
   +Did the girls attend the concert?

7. The hungry guests helped themselves/himself to the delicious meal.

8. The tired waitresses poured some coffee for themselves/herself after a hard day’s work.
   +Did the waitresses have some coffee?

9. The talented tailor made themselves/himself some beautiful clothes.

10. The careless scientist burned themselves/himself with the dangerous chemical.

11. The successful hunter cleaned themselves/himself off after walking through the woods.
   +Did the hunter walk through the woods?
12. The young mechanic considered himself/themselves to be very handsome.

13. The old lady gathered flowers for themselves/herself every morning during the spring.

-Did the lady plant flowers during the summer?

14. The busy housewife allowed herself/themselves a one-hour nap.

15. The tired nurse administered the injection herself/themselves without telling the doctor.

-Did the nurse tell the doctor about the injection?

16. The popular midwife taught herself/themselves how to dance.

+Was the midwife popular?

17. The lonely grandfather made himself/herself a cup of tea.

-Did the grandfather cook some soup?

18. The anxious cowboy prepared himself/herself for the performance.

19. The nervous son checked himself/herself in the large mirror.

+Did the son look in the mirror?

20. The young husband found himself/herself without a job or money.

+Was the husband poor?

21. The desperate boyfriend told himself/herself to forgive the girl.

22. The confused brother left himself/herself a note on the wall.

23. The stubborn gentleman did the work himself/herself during the meeting.

-Did the gentleman assign the work to someone else?
24. The proud king looked at himself/herself in the mirror in the living room.

25. The hungry waitress ordered herself/himself a big meal.

26. The successful woman congratulated herself/himself on the promotion.
+Was the woman promoted?

27. The wealthy queen built himself/herself a big castle in the forest.
-Did the queen build a small castle?

28. The new stepmother prepared herself/himself to meet the family.

29. The youthful grandmother planted a garden for herself/himself in the yard.
+Was the garden for the grandmother?

30. The brave girl told herself/himself not to worry.

31. The ambitious lady flew herself/himself to Paris for business.
-Did the lady go to Paris by train?

32. The new aunt introduced herself/himself to the guests.

33. The mother amused the father because he/she told funny jokes at dinner.

34. The landlady angered the policeman because he/she was cruel to animals.
+Did the landlady make the policeman angry?

35. The aunt scared the uncle because he/she got mad so easily.

36. The wife bored the husband because he/she told the same story over and over again.
+/-Did the husband tell the same story many times?

37. The congresswoman disappointed the chairman because he/she lost the election.
38. The gentleman frightened the hostess because he/she carried a bomb into the vehicle.

39. The princess apologized to the prince because he/she was sorry for the mistake.
+Did the princess apologize to the prince because of the mistake?

40. The brother annoyed the sister because he/she kept lying to people.
+Was the sister annoyed with the brother?

41. The lady inspired the actor because he/she volunteered at a homeless shelter.

42. The boy amazed the mother because he/she ran so fast in the race.
-Did the mother feel sad about the race?

43. The woman charmed the policeman because he/she was so adorable in the hat.
+Did the policeman like the hat?

44. The cowboy cheated the cowgirl because he/she wanted to win the game.

45. The uncle confessed to the policewoman because he/she felt guilty about the incident.
-Was the policewoman happy about the incident?

46. The grandson fascinated the grandmother because he/she was a wonderful singer.

47. The salesman deceived the girl because he/she wanted to sell the car.
-Was the car red?

48. The guy humiliated the lady because he/she told the story in public.

49. The boy admired the girl because he/she was so intelligent.
-Was the boy admired?
50. The nephew hated the niece because he/she cheated all the time.

51. The grandfather congratulated the granddaughter because he/she won the race.

52. The stepfather feared the daughter because he/she was such an aggressive person.

53. The widow envied the bachelor because he/she won the national lottery.

-Did the widow and the bachelor win the lottery?

54. The girl noticed the man because he/she was wearing such bright colors.

+Were the clothes brightly colored?

55. The actress praised the actor because he/she had given a good presentation yesterday.

-Was the presentation given last week?

56. The mother blamed the son because he/she missed the target in the game.

57. The granddaughter adored the father because he/she was the nicest person in the world.

+Was the father adored by the granddaughter?

58. The woman likes the boy because he/she is very honest.

59. The mother thanked the son because he/she had washed the car.

+Was the car clean?

60. The doorman assisted the lady because he/she was carrying a large suitcase.

61. The girl comforted the boy because he/she didn’t have the chance to go to college.

62. The policeman honored the grandmother because he/she deserved the award for saving people.
+-Did the policeman deserve the award?

63. The husband trusts the wife because he/she always tells the truth.

64. The stepfather despised the mother because he/she stole money from a church.

+Was money taken from a church?
APPENDIX B: OFFLINE TEST IN EXPERIMENT 1

Instructions: Complete the following sentences by circling the correct word in the parentheses.

Note: Items were scrambled in the actual test.

1. The lonely grandfather made_____(himself/herself) a cup of coffee.
2. The overweight cowboy denied____(himself/herself) a hamburger.
3. The successful man congratulated____(herself/himself) on the promotion.
4. The gracious wife introduced____(himself/herself) to the guests.
5. The infamous princess looked at ______(himself/herself) in the mirror.
6. The landlady worked____(himself/herself) into a frenzy.
7. The careless pedestrians found _____(themselves/himself) covered with mud.
8. The famous actresses prepared____(themselves/herself) to face the crowd.
9. The dirty soldiers cleaned____(themselves/himself) before going to bed.
10. The mother amused the father because____(he/she) told funny jokes.
11. The congressman disappointed the chairwoman because _____(he/she) lost the election.
12. The brother annoyed the sister because ______ (he/she) kept lying to people.
13. The widow envied the bachelor because _____ (he/she) won the national lottery.
14. The actress praised the actor because ______ (he/she) had given a good presentation.
15. The grandfather congratulated the granddaughter because_____ (he/she) won the reading competition.
APPENDIX C: LANGUAGE PROFILE

Language Profile

Subject #__________________

We are conducting research on second language acquisition. Please fill in the form as truthfully as possible. Information on this form is kept entirely confidential. Only your number code appears with the information you provide. By completing the questionnaire, you consent to our use of the information.

1. Date:__________
2. Age:___________ 3. Gender:_________   4. Place of birth:_____________
5. Do you have any hearing or vision problems? Yes______   No______
6. What is your mother tongue?_____________
7. What is your major?__________________________________________________
8. What is your TOEFL score (if you have one)?_________________
9. At what age did you start to learn English? ____________________
10. Years of residence in the US:_______Year(s)_______Month(s)
11. Years of formal education in total:____________________
12. Years of formal education in English speaking countries___________
13. Estimate your level of English on a scale of 1 (beginner) to 5 (advanced)
   Speaking  1  2  3  4  5  Listening  1  2  3  4  5
   Reading  1  2  3  4  5  Writing  1  2  3  4  5
14. List any other languages you speak besides English and your mother tongue________________

# APPENDIX D: RESULTS OF REACTION TIMES IN EXPERIMENT 1

<table>
<thead>
<tr>
<th>Native Speakers</th>
<th>Nonnative speakers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$t$    $df$  Sig.</td>
</tr>
<tr>
<td>Subject analysis (ungrammatical/implausible minus grammatical/plausible)</td>
<td></td>
</tr>
<tr>
<td>Number cue:</td>
<td></td>
</tr>
<tr>
<td>Position 1 NU1–NG1</td>
<td>-1.4</td>
</tr>
<tr>
<td>Position 2 NU2–NG2</td>
<td>-0.02</td>
</tr>
<tr>
<td>Position 3 NU3–NG3</td>
<td>.8</td>
</tr>
<tr>
<td>Position 4 NU4–NG4</td>
<td>4.1</td>
</tr>
<tr>
<td>Gender cue:</td>
<td></td>
</tr>
<tr>
<td>Position 1 GU1–GG1</td>
<td>1.2</td>
</tr>
<tr>
<td>Position 2 GU2–GG2</td>
<td>1.9</td>
</tr>
<tr>
<td>Position 3 GU3–GG3</td>
<td>7.5</td>
</tr>
<tr>
<td>Position 4 GU4–GG4</td>
<td>6.3</td>
</tr>
<tr>
<td>NP1-biased verb cue:</td>
<td></td>
</tr>
<tr>
<td>Position 1 OU1–OG1</td>
<td>-.59</td>
</tr>
<tr>
<td>Position 2 OU2–OG2</td>
<td>.03</td>
</tr>
<tr>
<td>Position 3 OU3–OG3</td>
<td>1.9</td>
</tr>
<tr>
<td>Position 4 OU4–OG4</td>
<td>2.6</td>
</tr>
<tr>
<td>NP2-biased verb cue:</td>
<td></td>
</tr>
<tr>
<td>Position 1 TU1–TG1</td>
<td>.59</td>
</tr>
<tr>
<td>Position 2 TU2–TG2</td>
<td>1.7</td>
</tr>
<tr>
<td>Position 3 TU3–TG3</td>
<td>3.2</td>
</tr>
<tr>
<td>Position 4 TU4–TG4</td>
<td>1.3</td>
</tr>
<tr>
<td>Item analysis(ungrammatical/implausible minus grammatical/plausible)</td>
<td></td>
</tr>
<tr>
<td>Number cue:</td>
<td></td>
</tr>
<tr>
<td>Position 1 NU1–NG1</td>
<td>-1.9</td>
</tr>
<tr>
<td>Position 2 NU2–NG2</td>
<td>-0.18</td>
</tr>
<tr>
<td>Position 3 NU3–NG3</td>
<td>5.7</td>
</tr>
<tr>
<td>Position 4 NU4–NG4</td>
<td>3.6</td>
</tr>
<tr>
<td>Gender cue:</td>
<td></td>
</tr>
<tr>
<td>Position 1 GU1–GG1</td>
<td>1.5</td>
</tr>
<tr>
<td>Position 2 GU2–GG2</td>
<td>2.9</td>
</tr>
<tr>
<td>Position 3 GU3–GG3</td>
<td>6.3</td>
</tr>
<tr>
<td>Position 4 GU4–GG4</td>
<td>6.0</td>
</tr>
</tbody>
</table>
NP1-biased verb cue:

<table>
<thead>
<tr>
<th>Position</th>
<th>OU1–OG1</th>
<th>31</th>
<th>.322</th>
<th>.94</th>
<th>31</th>
<th>.177</th>
</tr>
</thead>
<tbody>
<tr>
<td>Position</td>
<td>OU2–OG2</td>
<td>31</td>
<td>.467</td>
<td>2.2</td>
<td>31</td>
<td>.017*</td>
</tr>
<tr>
<td>Position</td>
<td>OU3–OG3</td>
<td>31</td>
<td>.040*</td>
<td>1.5</td>
<td>31</td>
<td>.077</td>
</tr>
<tr>
<td>Position</td>
<td>OU4–OG4</td>
<td>31</td>
<td>.023*</td>
<td>1.1</td>
<td>31</td>
<td>.135</td>
</tr>
</tbody>
</table>

NP2-biased verb cue:

<table>
<thead>
<tr>
<th>Position</th>
<th>TU1–TG1</th>
<th>31</th>
<th>.234</th>
<th>.46</th>
<th>31</th>
<th>.324</th>
</tr>
</thead>
<tbody>
<tr>
<td>Position</td>
<td>TU2–TG2</td>
<td>31</td>
<td>.067</td>
<td>1.5</td>
<td>31</td>
<td>.073</td>
</tr>
<tr>
<td>Position</td>
<td>TU3–TG3</td>
<td>31</td>
<td>.001*</td>
<td>1.6</td>
<td>31</td>
<td>.059</td>
</tr>
<tr>
<td>Position</td>
<td>TU4–TG4</td>
<td>31</td>
<td>.136</td>
<td>1.9</td>
<td>31</td>
<td>.033*</td>
</tr>
</tbody>
</table>

Note: The first letter indicates type of cues (N = number cues; G = gender cues; O=NP1-biased verb cues; T=NP2-biased verb cues); the second letter indicates version condition (U = ungrammatical/implausible; G = grammatical/plausible). And the number indicates the test position.
APPENDIX E: ONLINE TEST STIMULI IN EXPERIMENT 2

1. The son remembered that the father/mother introduced himself/herself at the meeting.
- Was the son introduced at the meeting?

2. The man said that the mother/grandson bought herself/himself a small gift from the shop.

3. John thought that Bill/Mary owed himself/herself another chance to solve the problem.

4. The cleaning lady knew that the old woman/man shot herself/himself in the hotel.
- Did the old woman/man shoot the cleaning lady in the hotel?

5. The nurse said that the soldier/policeman felt sorry for himself/herself because of the serious wound.
+/- Did the soldier/policeman feel happy?

6. The father thought that the daughter/son hated himself/herself because of the incident.

7. Rose said that Mike/Jill wrote about herself/himself in the letter.

8. Mary thought that Tom/Mike blamed herself/himself for being late for the train.
- Was the train late?

9. Mike believes that Helen/Tom trusts himself/herself to be able to get around New York.
10. Susan heard that Mary/John had bought herself/himself a new 10-speed bicycle. 
+/-Did Mary/John buy a new bicycle?

11. David could see that Bill/Sue was looking at himself/herself in the mirror.

12. Sam thinks that Jane/Tom dislikes herself/himself for being so impatient.
-Does Sam dislike Jane/Tom for being so impatient?

13. Alice knew that John/Mary understood herself/himself pretty well.

14. Bob said that Paul/Helen hit herself/himself with a long stick.

15. The waitress said the man/lady criticized himself/herself for being rude at the restaurant.
+/-Was the man/lady rude at the restaurant?

16. The woman remembered that the daughter/son brought herself/himself a beautiful bouquet of flowers.
+/-Did the daughter/son buy some beautiful flowers?

17. The little girl was happy that the father/mother bought herself/himself a nice toy last week.

18. Mary angrily told me that the boy/girl had spilled a lot of paint on herself/himself the other day.
+Did the boy/girl spill some paint the other day?

19. The little boy was angry that the father/mother hit himself/herself last week at home.

20. The famous actress heard that the popular host/hostess talked about
himself/herself on TV last night.

-Did the famous actress watch the show on TV last night?

21. The lady said unhappily that the brother/sister often does not trust herself/himself because of the nasty gossip.

22. The cashier was worried that Mrs. Smith/Mr. Smith wrote herself/himself a bad check.

23. The student was angry that Rose/Jack gave himself/herself so much homework.

+/-Did Rose/Jack have so much homework?

24. The boy was excited that the mother/father bought himself/herself a birthday gift.

25. George felt embarrassed that the woman/man found himself/herself staring at an attractive nurse.

-Was George embarrassed about staring at an attractive nurse?

26. The talkative uncle was sad that the father/mother kept himself/herself from talking to friends.

27. The nun was mad that the priest/priestess kept looking at himself/herself after the sermon.

-Was the nun happy with the priest/priestess?

28. The husband was embarrassed that the lady/guy talked himself/herself into believing the bizarre story.

29. The hungry boy was very happy that the grandfather/grandmother brought himself/herself a chocolate cake.
30. The little girl with a bad cold was told that the man/woman would take care of himself/herself during the journey.

+Would the man/woman look after the little girl during the journey?

31. The grandson was disappointed that the father/mother didn’t buy himself/herself a new toy car from the shop.

+Did the grandson feel disappointed about the new toy car?

32. The new chairwoman was frustrated that the new actor/actress didn’t trust himself/herself to finish the big project.

-Did the new actor/actress distrust the new chairwoman?
APPENDIX F: OFFLINE TEST IN EXPERIMENTS 2 AND 3

Instruction: Answer each question in the following about the sentences you read.

Note: Items were scrambled in the actual test.

1. The little boy was angry that the father always hit himself.
   Can “himself” refer to the little boy? Yes  No
   Can “himself” refer to the father? Yes  No
   Can “himself” refer to somebody else? Yes  No

2. The hungry girl was happy that the mother bought herself an ice cream.
   Can “herself” refer to the hungry girl? Yes  No
   Can “herself” refer to the mother? Yes  No
   Can “herself” refer to somebody else? Yes  No

3. The customer was upset that the waiter spilled some soup on himself.
   Can “himself” refer to the customer? Yes  No
   Can “himself” refer to the waiter? Yes  No
   Can “himself” refer to somebody else? Yes  No

4. The man with the little boy bought himself a new house in the city.
   Can “himself” refer to the man? Yes  No
   Can “himself” refer to the little boy? Yes  No
   Can “himself” refer to somebody else? Yes  No

5. Jill heard that the famous actress talked about herself on TV.
Can “herself” refer to Jill? Yes  No
Can “herself” refer to the famous actress? Yes  No
Can “herself” refer to somebody else? Yes  No

6. Mike said that the little boy was dressing himself upstairs.
Can “himself” refer to Mike? Yes  No
Can “himself” refer to the little boy? Yes  No
Can “himself” refer to somebody else? Yes  No

7. The patient beside the doctor can look after himself next week.
Can “himself” refer to the patient? Yes  No
Can “himself” refer to the doctor? Yes  No
Can “himself” refer to somebody else? Yes  No

8. Jordon remembered that the popular singer wrote about himself in the autography.
Can “himself” refer to Jordon? Yes  No
Can “himself” refer to the popular singer? Yes  No
Can “himself” refer to somebody else? Yes No

9. The boy scout next to Bill introduced himself at the meeting.
Can “himself” refer to the boy? Yes  No
Can “himself” refer to Bill? Yes  No
Can “himself” refer to somebody else? Yes  No

10. Rose said that Jill voted for herself.
Can “herself” refer to Rose? Yes  No
Can “herself” refer to Jill? Yes  No
Can “herself” refer to somebody else? Yes  No

11. Jack says that Peter talked about himself.
Can “himself” refer to Jack? Yes  No
Can “himself” refer to Peter? Yes  No
Can “himself” refer to somebody else? Yes  No

12. The nurse said that the old woman killed herself.
Can “herself” refer to the nurse? Yes  No
Can “herself” refer to the old woman? Yes  No
Can “herself” refer to somebody else? Yes  No
## APPENDIX G: RESULTS OF REACTION TIMES IN EXPERIMENT 2

<table>
<thead>
<tr>
<th>Native Speakers</th>
<th>Nonnative speakers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( t )</td>
</tr>
<tr>
<td>Subject analysis</td>
<td></td>
</tr>
<tr>
<td>Neutral sentences:</td>
<td></td>
</tr>
<tr>
<td>Both match vs. no match:</td>
<td></td>
</tr>
<tr>
<td>Position 1 BU1–NU1</td>
<td>.61</td>
</tr>
<tr>
<td>Position 2 BU2–NU2</td>
<td>-.07</td>
</tr>
<tr>
<td>Position 3 BU3–NU3</td>
<td>-6.4</td>
</tr>
<tr>
<td>Position 4 BU4–NU4</td>
<td>-2.2</td>
</tr>
<tr>
<td>Only NP1 match vs. only NP2 match:</td>
<td></td>
</tr>
<tr>
<td>Position 1 OU1–TU1</td>
<td>-.03</td>
</tr>
<tr>
<td>Position 2 OU2–TU2</td>
<td>-.34</td>
</tr>
<tr>
<td>Position 3 OU3–TU3</td>
<td>3.3</td>
</tr>
<tr>
<td>Position 4 OU4–TU4</td>
<td>1.9</td>
</tr>
<tr>
<td>Biased sentences:</td>
<td></td>
</tr>
<tr>
<td>Both match vs. no match:</td>
<td></td>
</tr>
<tr>
<td>Position 1 BG1–NG1</td>
<td>-.37</td>
</tr>
<tr>
<td>Position 2 BG2–NG2</td>
<td>-1.2</td>
</tr>
<tr>
<td>Position 3 BG3–NG3</td>
<td>-3.7</td>
</tr>
<tr>
<td>Position 4 BG4–NG4</td>
<td>-2.7</td>
</tr>
<tr>
<td>Only NP1 match vs. only NP2 match:</td>
<td></td>
</tr>
<tr>
<td>Position 1 OG1–TG1</td>
<td>-.33</td>
</tr>
<tr>
<td>Position 2 OG2–TG2</td>
<td>-.78</td>
</tr>
<tr>
<td>Position 3 OG3–TG3</td>
<td>4.9</td>
</tr>
<tr>
<td>Position 4 OG4–TG4</td>
<td>1.2</td>
</tr>
</tbody>
</table>

Item analysis

**Neutral sentences:**
Both match vs. No match:
<p>| Position 1 BU1–NU1 | .51 | 47 | .306 | -.16 | 47 | .437 |
| Position 2 BU2–NU2 | .67 | 47 | .252 | .20 | 47 | .422 |
| Position 3 BU3–NU3 | -4.9 | 47 | .000* | -1.1 | 47 | .150 |
| Position 4 BU4–NU4 | -2.0 | 47 | .026* | -2.0 | 47 | .027* |</p>
<table>
<thead>
<tr>
<th>Position</th>
<th>OU1–TU1</th>
<th>47</th>
<th>.434</th>
<th>.42</th>
<th>47</th>
<th>.340</th>
</tr>
</thead>
<tbody>
<tr>
<td>Position 2</td>
<td>OU2–TU2</td>
<td>-1.3</td>
<td>.107</td>
<td>1.3</td>
<td>47</td>
<td>.108</td>
</tr>
<tr>
<td>Position 3</td>
<td>OU3–TU3</td>
<td>2.2</td>
<td>.014*</td>
<td>2.3</td>
<td>47</td>
<td>.015*</td>
</tr>
<tr>
<td>Position 4</td>
<td>OU4–TU4</td>
<td>2.9</td>
<td>.003*</td>
<td>1.6</td>
<td>47</td>
<td>.061</td>
</tr>
</tbody>
</table>

**Biased sentences:**

Both match vs. No match:

<table>
<thead>
<tr>
<th>Position 1</th>
<th>BG1–NG1</th>
<th>47</th>
<th>.35</th>
<th>-.05</th>
<th>47</th>
<th>.480</th>
</tr>
</thead>
<tbody>
<tr>
<td>Position 2</td>
<td>BG2–NG2</td>
<td>47</td>
<td>.02</td>
<td>-.46</td>
<td>47</td>
<td>.324</td>
</tr>
<tr>
<td>Position 3</td>
<td>BG3–NG3</td>
<td>47</td>
<td>-.4.2</td>
<td>-.2.5</td>
<td>47</td>
<td>.008*</td>
</tr>
<tr>
<td>Position 4</td>
<td>BG4–NG4</td>
<td>47</td>
<td>-.2.1</td>
<td>-.2.0</td>
<td>47</td>
<td>.028*</td>
</tr>
</tbody>
</table>

Only NP1 match vs. only NP2 match:

<table>
<thead>
<tr>
<th>Position 1</th>
<th>OG1–TG1</th>
<th>47</th>
<th>-.21</th>
<th>-.76</th>
<th>47</th>
<th>.225</th>
</tr>
</thead>
<tbody>
<tr>
<td>Position 2</td>
<td>OG2–TG2</td>
<td>47</td>
<td>-1.2</td>
<td>-.86</td>
<td>47</td>
<td>.196</td>
</tr>
<tr>
<td>Position 3</td>
<td>OG3–TG3</td>
<td>47</td>
<td>4.5</td>
<td>1.6</td>
<td>47</td>
<td>.054*</td>
</tr>
<tr>
<td>Position 4</td>
<td>OG4–TG4</td>
<td>47</td>
<td>1.3</td>
<td>.98</td>
<td>47</td>
<td>.166</td>
</tr>
</tbody>
</table>

Note: The first letter indicates type of conditions (B = both match; N = no match; O = only NP1 match; T = only NP2 match); the second letter indicates sentence categories (U = neutral; G = biased); the number indicates the test position. Thus, BU1–NU1 means the comparison of RT means on the “both match” and “no match” sentences involving the structural cue at the first position.
APPENDIX H: OFFLINE TEST IN EXPERIMENT 3

I. Answer each of the questions below (Version A; Subject ID_______)

1. Rose said that Mary voted for herself.
   Can “herself” refer to Rose? Yes   No
   Can “herself” refer to Mary? Yes   No
   Can “herself” refer to somebody else? Yes   No

2. The little boy was angry that the father always hit himself.
   Can “himself” refer to the little boy? Yes   No
   Can “himself” refer to the father? Yes   No
   Can “himself” refer to somebody else? Yes   No

3. Mary heard that the famous actress talked about herself on TV.
   Can “herself” refer to Mary? Yes   No
   Can “herself” refer to the famous actress? Yes   No
   Can “herself” refer to somebody else? Yes   No

4. The man next to the little boy bought himself a new toy from the store.
   Can “himself” refer to the man? Yes   No
   Can “himself” refer to the little boy? Yes   No
   Can “himself” refer to somebody else? Yes   No

5. Jane said that the little girl was dressing herself upstairs.
   Can “herself” refer to Jane? Yes   No
Can “herself” refer to the little girl? Yes No

Can “herself” refer to somebody else? Yes No

6. Jordon remembered that the popular singer wrote about himself in the autography.

Can “himself” refer to Jordon? Yes No

Can “himself” refer to the popular singer? Yes No

Can “himself” refer to somebody else? Yes No

7. The hungry girl was happy that the mother bought herself an ice cream.

Can “herself” refer to the hungry girl? Yes No

Can “herself” refer to the mother? Yes No

Can “herself” refer to somebody else? Yes No

8. The lady next to Kate introduced herself at the meeting.

Can “herself” refer to the lady? Yes No

Can “herself” refer to Kate? Yes No

Can “herself” refer to somebody else? Yes No

9. Jack says that Peter talked about himself.

Can “himself” refer to Jack? Yes No

Can “himself” refer to Peter? Yes No

Can “himself” refer to somebody else? Yes No

10. The customer was upset that the waiter spilled some soup on himself.

Can “himself” refer to the customer? Yes No

Can “himself” refer to the waiter? Yes No
Can “himself” refer to somebody else? Yes No

11. The patient beside the doctor can look after himself next week.

Can “himself” refer to the patient? Yes No
Can “himself” refer to the doctor? Yes No
Can “himself” refer to somebody else? Yes No

12. The lady said that the old woman blamed herself for being late for the train.

Can “herself” refer to the lady? Yes No
Can “herself” refer to the old woman? Yes No
Can “herself” refer to somebody else? Yes No

II. Read the story and then complete the last sentence with a pronoun or reflexive.

1. Once a week, Mike used to visit an old woman. On Mike’s last visit he saw the old woman point a gun at her head and shoot. The old woman died instantly.

Mike knew that the old woman shot_________.
A. him        B. himself  C. her        D. herself

2. Johnny and a little boy were playing with matches. Johnny lit a match and then dropped it on the little boy’s leg. The little boy went screaming to his father and told him what had happened.

The little boy said that Johnny had burned_________.
A. him        B. himself  C. her        D. herself
3. Bill was going to a party. A very famous male actor was going to attend the party. Bill was too shy to speak to the actor, so he hoped the actor would speak to him instead.

Bill hoped that the famous actor would introduce______.
A. him        B. himself  C. her        D. herself

4. Killer Harry was free again. Bill was very scared. Bill called a policeman so the policeman could guard him and make sure he was safe from Killer Harry.

Bill asked the policeman to protect______.
A. him        B. himself  C. her        D. herself

I. Answer each of the questions below (Version B; Subject ID______)

1. Bill said that Peter criticized himself for being rude at the restaurant.
   Can “himself” refer to Bill? Yes   No
   Can “himself” refer to Peter? Yes   No
   Can “himself” refer to somebody else? Yes   No

2. The patient was worried that the nurse would spill some medicine on herself.
   Can “herself” refer to the patient? Yes   No
   Can “herself” refer to the nurse? Yes   No
   Can “herself” refer to somebody else? Yes   No

3. Mary believed that the famous actress talked about herself on TV.
   Can “herself” refer to Mary? Yes   No
Can “herself” refer to the famous actress? Yes  No
Can “herself” refer to somebody else? Yes  No

4. The son remembered that the father introduced himself at the party.
Can “himself” refer to the son? Yes  No
Can “himself” refer to the father? Yes  No
Can “himself” refer to somebody else? Yes  No

5. The nurse next to Helen can wake up herself the next morning.
Can “herself” refer to the nurse? Yes  No
Can “herself” refer to Helen? Yes  No
Can “herself” refer to somebody else? Yes  No

6. Bill said that the actor was preparing himself upstairs for the performance.
Can “himself” refer to Bill? Yes  No
Can “himself” refer to the actor? Yes  No
Can “himself” refer to somebody else? Yes  No

7. The mother said that the daughter would buy herself a new toy.
Can “herself” refer to the mother? Yes  No
Can “herself” refer to the daughter? Yes  No
Can “herself” refer to somebody else? Yes  No

8. The tourist was angry that the tour guide laughed at himself.
Can “himself” refer to the tourist? Yes  No
Can “himself” refer to the tour guide? Yes  No
Can “himself” refer to somebody else? Yes No

9. The thirsty boy behind the father bought himself a bottle of water.

Can “himself” refer to the thirsty boy? Yes No
Can “himself” refer to the father? Yes No
Can “himself” refer to somebody else? Yes No

10. Mike was upset that the son made fun of himself in the classroom.

Can “himself” refer to Mike? Yes No
Can “himself” refer to the son? Yes No
Can “himself” refer to somebody else? Yes No

11. The lady beside Mary made herself some new clothes.

Can “herself” refer to the lady? Yes No
Can “herself” refer to Mary? Yes No
Can “herself” refer to somebody else? Yes No

12. The mother said that Susan prepared herself a meal.

Can “herself” refer to the mother? Yes No
Can “herself” refer to Susan? Yes No
Can “herself” refer to somebody else? Yes No

II. Read the story and then complete the last sentence with a pronoun or reflexive.

1. After three years in the war, the soldier finally went crazy and jumped out of a
window. He died instantly. The doctor had to tell the soldier’s family the sad
news.

The doctor said that the soldier killed_____.

A. him        B. himself  C. her       D. herself

2. Susan and her friend were sewing. They were careless and left some pins on the
floor. Susan was not wearing shoes and she stepped on a pin. Susan started to
shout and cry. Susan’s friend could see the blood on Susan’s foot.

The friend realized that Susan hurt_____.

A. him        B. himself  C. her       D. herself

3. Mary used to have a problem. Every time she met someone she got nervous and
forgot her own name. Mary was going to a party at a friend’s house. Mary hoped
her friend would help Mary meet people by telling them Mary’s name.

Mary hoped that her friend would introduce_____.

A. him        B. himself  C. her       D. herself

4. Johnny and his father were going for a walk. It started to rain. Johnny had a cold,
so his father gave him a jacket and told him to put it over his head. Johnny felt
warm under the jacket.

His father told Johnny to cover______ with a jacket.

A. him        B. himself  C. her       D. herself
APPENDIX I: COMPUTER-DELIVERED PROCESSING INSTRUCTION IN EXPERIMENT 3

A. Select the picture that goes with the sentence.

1. Mr. Green explained that Mr. Blue cut himself.

2 Mr. Green said that Mr. Blue cut him. (Same pictures as above)

3. Mr. Green said angrily that Mr. Blue hit himself.
4. The man next to Mr. Green looked at himself in the mirror.
5. Mr. Blue said that Mr. Red painted himself.

B. Read each sentence, and then indicate who is performing the action by answering each question.

1. Lee remembered that Andrew introduced himself at the meeting.

Whom did Andrew introduce?

a. Lee b. Andrew c. somebody else d. either Lee or Andrew

2. Mary said that Jill understands herself.

Whom did Jack understand?

a. Mary b. Jill c. either Mary or Jill d. somebody else
3. The hungry boy Jack was happy that John brought himself some cakes.

Whom did John bring cakes for?

a. Jack   b. John   c. either Jack or John   d. somebody else

4. Mrs. Brown was angry that Mrs. Black blamed her for the wrong decision.

Whom did Mrs. Black blame?

a. Mrs. Brown   b. Mrs. Black   c. either Mrs. Brown or Mrs. Black   d. somebody else

5. The lady beside June hit herself.

Whom did the lady hit?

a. the lady   b. June   c. Either the lady or June   d. somebody else


Whom did Bill blame?

a. Bill   b. Tom   c. either Bill or Tom   d. somebody else

7. The depressed woman said that the mother should leave her alone.

Who should be left alone?

a. the woman   b. the mother   c. either the woman or the mother   d. somebody else.

C. Select a female classmate or relative or friend of yours (mother, sister, aunt, niece, etc.) and write her name below. Which of the following statements is likely true to her?

Name_______
1. She says that her father will buy himself a new car.  True____   Not True___

2. She hopes that her father would buy himself a house in Hawaii. True___   Not True___

3. She once said that her grandmother could take care of herself. True____   Not True___

4. Her grandmother once asked her to bring herself cookies to school. True____
   Not True___

5. She says that if she can speak English well her mother will be proud of herself.
   True____   Not True___

Select a male classmate or relative or friend and do the same.

Name________________

1. He says that his father will buy himself a new car.  True____   Not True___

2. He hopes that his mother would buy herself a house in Hawaii. True___   Not True___

3. He once said that his grandmother could take care of herself. True____   Not True___

4. His grandmother once asked him to bring himself cookies to school. True____
   Not True___

5. He says that if he can speak English well his father will be proud of himself.
   True____   Not True___
D. Select the best way to complete each sentence.

1. The little girl was happy that her father bought ________a nice toy.
   a. herself  b. himself  c. either a or b

2. The cashier was angry that the woman wrote ______ a fake check.
   a. himself  b. herself  c. either a or b

3. The waitress behind the man calmed ____ with a glass of wine.
   a. himself  b. herself  c. either a or b

4. The little girl in front of the policeman protected ____ from the poisonous plant.
   a. himself  b. herself  c. either a or b

5. The famous actor heard that the popular hostess talked about ______ on TV last week.
   a. himself  b. herself  c. either a or b

E. Listen to each sentence, and then indicate who is performing the action by answering each question.

1. The hungry housewife was grateful that the husband cooked himself during the weekend.
   
   Who is “himself” in the statement?
   
   a. the housewife  b. the husband  c. either the housewife or the husband  d. somebody else

2. Mr. Brown said that Mr. Green sprayed himself with insect repellent.
Who is “himself” in the statement?

a. Mr. Brown  b. Mr. Green  c. either Mr. Brown or Mr. Green  d. somebody else

3. Mr. Black was angry that Mr. Green did not trust him.

Who is “him”?

a. Mr. Black  b. Mr. Green  c. either Mr. Black or Mr. Green  d. somebody else

4. Mr. Black knew that Mr. Brown killed himself.

Who is “himself”?

a. Mr. Black  b. Mr. Green  c. either Mr. Black or Mr. Green  d. somebody else

5. Mrs. Black was outraged that Mrs. Green locked herself in the car.

Who is “herself”?

a. Mrs. Black  b. Mrs. Green  c. either Mrs. Black or Mrs. Green  d. somebody else

6. Mr. Brown dreamed that Mr. Green shot himself.

Who is “himself”?

a. Mr. Brown  b. Mr. Green  c. either Mr. Brown or Mr. Green  d. somebody else

F. Read the story and then complete the sentence according to the content of the story.
1. Bill is a student. There was a new teacher in his class today. During class, the teacher asked Bill’s name and Bill’s hometown. Bill told the teacher that he was born in New York.

The teacher said that Bill introduced ______.

2. Susan wanted a job in the hospital. A nurse interviewed Susan for the job. The nurse asked Susan about her experience, her education and whether she got on well with people. The nurse told the director about the interview.

The nurse said that Susan talked about ______ in the interview.

3. A young boy was looking at one of Mr. Robins’s guns. The young boy accidentally pulled the trigger and the gun fired. Unfortunately, the bullet hit Mr. Robins in the arm.

Mr. Robins realized that the boy shot ______ accidentally.

4. Annie wanted to get her friend into trouble. Annie went to the closet and suggested that her friend lock the door from the outside. When her friend followed her advice, Annie started to cry. When Annie got out finally, she told her mother about it.

Annie said that her friend locked ______ in the closet.

G. Read the passage. Then answer the questions that follow.

Parent: Can my son drive (1) himself to games, Director?

Director: The university provides transportation to all games and students are
expected to travel together as a team. Under special circumstances, the team director may allow a student to drive (2) *himself* only after a parent has made those arrangements with the director in advance.

**Student**: The team director says that a student cannot drive (3) *himself* to games except under special circumstances.

1. Who does the underlined (1) *himself* refer to?
   a. director   b. son   c. either a or b   d. somebody else.

2. Who does the underlined (2) *himself* refer to?
   a. the team director   b. a student   c. either a or b   d. somebody else.

3. Who does the underlined (3) *himself* refer to?
   a. the team director   b. a student   c. either a or b   d. somebody else.

4. In the underlined (3) *himself*, can we replace himself with herself?
   Yes __   No___
APPENDIX J: EXIT SURVEY IN EXPERIMENT 3

1. This lesson is an interactive computer-based lesson. How enjoyable was the learning process compared to paper-pencil classroom activities?
   a. much more enjoyable  b. somewhat more enjoyable  c. about as enjoyable  d. somewhat less enjoyable  e. much less enjoyable

2. Indicate how strongly you agree or disagree with the following statement:
   Because this activity is presented in an interactive computer-based format, I was better able to understand the ideas and concepts taught in it.
   a. strongly agree  b. agree  c. disagree  d. strongly disagree

3. Indicate how strongly you agree or disagree with the following statement:
   Because this activity is presented in an interactive computer-based format, I was at a disadvantage because I do not have adequate computer skills.
   a. strongly agree  b. agree  c. disagree  d. strongly disagree

4. Indicate how useful the feedback with explicit explanation of grammar rules in this lesson was:
   a. I did not get grammar rules.
   b. extremely useful  c. very useful  d. useful  e. somewhat useful  f. not useful

5. Notice that in this lesson, you did not produce or speak pronouns.
   Indicate how strongly you agree or disagree with the following statement:
   In order to learn reflexive pronouns better, I would like to have chances to produce or
speak reflexive pronouns during the lesson.

a. strongly agree  b. agree  c. disagree   d. strongly disagree
### APPENDIX K: RESULTS OF REACTION TIMES IN EXPERIMENT 3

<table>
<thead>
<tr>
<th></th>
<th>Pretest</th>
<th></th>
<th>Posttest</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>t</td>
<td>df</td>
<td>Sig.</td>
<td>t</td>
</tr>
<tr>
<td>Subject analysis</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neutral sentences:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Both match vs. no match:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Position 1 BU1–NU1</td>
<td>-.44</td>
<td>27</td>
<td>.332</td>
<td>.16</td>
</tr>
<tr>
<td>Position 2 BU2–NU2</td>
<td>-.75</td>
<td>27</td>
<td>.229</td>
<td>-.58</td>
</tr>
<tr>
<td>Position 3 BU3–NU3</td>
<td>-.20</td>
<td>27</td>
<td>.030*</td>
<td>-2.5</td>
</tr>
<tr>
<td>Position 4 BU4–NU4</td>
<td>-1.6</td>
<td>27</td>
<td>.062</td>
<td>-.39</td>
</tr>
<tr>
<td>Only NP1 match vs. only NP2 match:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Position 1 OU1–TU1</td>
<td>.05</td>
<td>27</td>
<td>.481</td>
<td>1.3</td>
</tr>
<tr>
<td>Position 2 OU2–TU2</td>
<td>.31</td>
<td>27</td>
<td>.379</td>
<td>-.82</td>
</tr>
<tr>
<td>Position 3 OU3–TU3</td>
<td>1.6</td>
<td>27</td>
<td>.057</td>
<td>2.4</td>
</tr>
<tr>
<td>Position 4 OU4–TU4</td>
<td>-.44</td>
<td>27</td>
<td>.332</td>
<td>1.6</td>
</tr>
<tr>
<td>Biased sentences:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Both match vs. no match:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Position 1 BG1–NG1</td>
<td>1.7</td>
<td>27</td>
<td>.051*</td>
<td>.19</td>
</tr>
<tr>
<td>Position 2 BG2–NG2</td>
<td>-.67</td>
<td>27</td>
<td>.254</td>
<td>-1.2</td>
</tr>
<tr>
<td>Position 3 BG3–NG3</td>
<td>-2.2</td>
<td>27</td>
<td>.018*</td>
<td>-3.4</td>
</tr>
<tr>
<td>Position 4 BG4–NG4</td>
<td>.13</td>
<td>27</td>
<td>.451</td>
<td>-3.4</td>
</tr>
<tr>
<td>Only NP1 match vs. only NP2 match:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Position 1 OG1–TG1</td>
<td>.79</td>
<td>27</td>
<td>.220</td>
<td>.11</td>
</tr>
<tr>
<td>Position 2 OG2–TG2</td>
<td>-.51</td>
<td>27</td>
<td>.309</td>
<td>-.21</td>
</tr>
<tr>
<td>Position 3 OG3–TG3</td>
<td>-.78</td>
<td>27</td>
<td>.222</td>
<td>.55</td>
</tr>
<tr>
<td>Position 4 OG4–TG4</td>
<td>-.22</td>
<td>27</td>
<td>.414</td>
<td>1.9</td>
</tr>
<tr>
<td>Item analysis</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neutral sentences:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Both match vs. No match:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Position 1 BU1–NU1</td>
<td>-.20</td>
<td>47</td>
<td>.421</td>
<td>.08</td>
</tr>
<tr>
<td>Position 2 BU2–NU2</td>
<td>-.34</td>
<td>47</td>
<td>.368</td>
<td>.36</td>
</tr>
<tr>
<td>Position 3 BU3–NU3</td>
<td>-2.0</td>
<td>47</td>
<td>.028*</td>
<td>-1.9</td>
</tr>
<tr>
<td>Position 4 BU4–NU4</td>
<td>-1.1</td>
<td>47</td>
<td>.138</td>
<td>-.16</td>
</tr>
<tr>
<td>Only NP1 match vs. only NP2 match:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: * indicates significance at the 0.05 level.
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.02</td>
<td>2</td>
<td>.83</td>
<td>3</td>
<td>1.2</td>
<td>4</td>
<td>-.24</td>
</tr>
<tr>
<td>47</td>
<td>.491</td>
<td>47</td>
<td>.207</td>
<td>47</td>
<td>.127</td>
<td>47</td>
<td>.406</td>
</tr>
<tr>
<td>.84</td>
<td>47</td>
<td>.203</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Biased sentences:**

Both match vs. No match:

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.4</td>
<td>2</td>
<td>.06</td>
<td>3</td>
<td>-1.6</td>
<td>4</td>
<td>-.15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>47</td>
<td>.083</td>
<td>47</td>
<td>.475</td>
<td>47</td>
<td>.054*</td>
<td>47</td>
<td>.441</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-.12</td>
<td>47</td>
<td>.451</td>
<td>.14</td>
<td>47</td>
<td>.443</td>
<td>-3.0</td>
<td>47</td>
<td>.003*</td>
<td></td>
</tr>
</tbody>
</table>

Only NP1 match vs. only NP2 match:

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.54</td>
<td>2</td>
<td>-.23</td>
<td>3</td>
<td>-.65</td>
<td>4</td>
<td>-.14</td>
</tr>
<tr>
<td>47</td>
<td>.298</td>
<td>47</td>
<td>.409</td>
<td>47</td>
<td>.261</td>
<td>47</td>
<td>.446</td>
</tr>
<tr>
<td>.53</td>
<td>47</td>
<td>.301</td>
<td>-.54</td>
<td>47</td>
<td>.297</td>
<td>.32</td>
<td>47</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: The first letter indicates type of conditions (B = both match; N = no match; O=only NP1 match; T=only NP2 match); the second letter indicates sentence categories (U = neutral; G = biased); the number indicates the test position. Thus, BU1–NU1 means the comparison of RT means on the “both match” and “no match” sentences involving the structural cue at the first position.


Doughty, C. J. (2004). Commentary: When PI is focus on form it is very very good, but when it is focus on forms…. In B.VanPatten (Ed.), *Processing instruction: Theory, research, and commentary* (pp. 257-270) Mahwah, NJ: Erlbaum.


