ABSTRACT

Title of Document: GRAMMATICAL GENDER REPRESENTATION AND PROCESSING IN ADVANCED SECOND LANGUAGE LEARNERS OF FRENCH

Karen L. Vatz, Ph.D., 2009

Directed By: Dr. Robert DeKeyser, Second Language Acquisition

One of the most difficult challenges of learning French as a foreign language is mastering the gender system. Although there are theories that account for how French native speakers (NSs) master their gender system, it is not fully understood why second language (L2) learners are unable to do the same. The goal of the present study was to investigate this difference in ability between French NSs and non-native speakers (NNSs), specifically, how L2 learners of French store grammatical gender knowledge, and how their storage system relates to processing of grammatical gender in terms of the ability to realize accurate gender agreement throughout a sentence.

First, a gender priming task investigated whether advanced L2 learners have developed a gender-nodal system in which gender information is stored as an inherent property of a noun. Second, an online grammaticality judgment task addressed L2 learners' gender agreement ability during processing, while taking into account (a) the role of gender cues available to the participant, and (b) non-linguistic processing constraints such as working memory (WM) through manipulating the distance of an adjective from the noun with which it must agree. In order to investigate the role of a learner's native language (L1) in gender representation and processing, participants included learners of French from three L1 groups: Spanish, whose gender system is congruent to that of French; Dutch, whose gender system is incongruent to that of French; and English, whose gender system is minimal, relative to French. A group of NS controls also participated.

Results from the gender priming task indicate that the NNSs in the current study have not developed a native-like gender-nodal system, regardless of L1-L2 gender-system similarity. At-chance accuracy on the grammaticality judgment task indicates L2 gender agreement is far from native-like, even for advanced learners. Whereas the presence of gender cues was beneficial, neither WM nor L1-L2 similarity facilitated performance. The results from this study confirm previous findings on the difficulty of L2 gender agreement, and shed light on the nature of L2 gender representation as a possible explanation for this processing difficulty.

GRAMMATICAL GENDER REPRESENTATION AND PROCESSING IN ADVANCED SECOND LANGUAGE LEARNERS OF FRENCH

By

Karen L. Vatz

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2009

Advisory Committee:
Professor Robert DeKeyser, Chair
Dr. Erica B. Michael
Professor Catherine Doughty
Professor Mary Ellen Scullen
Dean’s Representative: Professor Rochelle Newman

Acknowledgements

Many people contributed to the completion of this dissertation, and it is a pleasure to thank all who made it possible. First, I would like to acknowledge funding from the University of Maryland Graduate School Ann G. Wylie Dissertation Fellowship, the Center for Advanced Study of Language Dissertation Fellowship, and the Language Learning Dissertation Grant.

I am very grateful to my doctoral committee for their wisdom and willingness to work with me: Erica Michael, Robert DeKeyser, Cathy Doughty, Mel Scullen, and Rochelle Newman. Erica, it has been a true privilege working with you. Thank you for all you have taught me and your commitment to this project. You have been a wonderful teacher and mentor and I will miss sitting in your office poring over data on Friday afternoons. Your calm, rational, and insightful approach to research made dissertating a joy. I hope to continue learning from you, as I am not ready for that aspect to end. Cathy, thank you for your investment in me over the last five years. Your input in the early stages of this project helped to shape it, and your thorough feedback along the way has been invaluable. Robert, in addition to your input in the research design and your help with the French, Spanish, and Dutch materials and translations, I am forever grateful for the many hours you have spent guiding us through the paperwork. Without your help, we would not have gotten to the end. Mel, you brought me to Maryland and were the first to suggest I take a course in psycholinguistics; I would not have taken this path without your early guidance. Rochelle, thank you for your input on the analyses; this research is more complete because of your contribution.

I would especially like to thank Joy Darling Johnston. Without your help and meticulous work, data collection would not have been possible. I hope our paths cross again; it was a pleasure working with you.

There are many people who helped with participant recruitment: Peter Lauwers, Aurélie Welcomme, Marie-Eve Michot, Mathea Simons, Aline Godfroid, Gisela Granena, Professors Brami and Letzter, Anne Bovaird, Sally Morris, and Dominique Meys. Thank you all for helping me find so many Spanish, English, and Dutch learners of French.

A special thanks to Jared Linck, Allison Blodgett, and Barb Forsyth for your help with statistical questions.

Thank you to my family for all the support along the way. I could not have done it without your help with childcare, editing drafts, and encouragement over the years.

The biggest thanks of all goes to David. This dissertation is yours as much as mine. Tech support, formatting, late night cookies, patience, encouragement, calm, perspective… I wouldn't have even tried without you. Thank you for everything. Now it's your turn!

And finally, to Maia and Adeline who arrived along the way. Maia, you remind me every day about the importance and joy of keeping a balance, and Addie, your arrival was the driving force that got me through to the end.

Merci à tous.

Table of Contents

Acknowledgements ii
Table of Contents iv
List of Tables vii
List of Figures ix
Chapter 1: Introduction 1
Chapter 2: French Gender System and L1 Gender Acquisition 5
 2.1 French Gender System 5
  2.1.1 Phonological Rules 7
  2.1.2 Morphological Rules 10
  2.1.3 Semantic Rules 12
 2.2 L1 Gender Acquisition 13
  2.2.1 L1 Gender Acquisition Data 14
  2.2.2 L1 Gender Assignment Data in adults 19
Chapter 3: Model of Grammatical Gender Processing 21
 3.1 Gender Processing within Levelt's Model of Speech Production 21
 3.2 L1 Gender Processing Studies 24
  3.2.1 Empirical support for Levelt's model: Physiological studies 24
  3.2.2 Empirical support for Levelt's model: Natural speech production 26
  3.2.3 Empirical support for Levelt's model: Behavioral studies 28
 3.3 Relationship Between Lemma and Gender Information 34
Chapter 4: L2 Gender Assignment and Agreement Ability 38
 4.1 Evidence for Non-native-like L2 Gender Agreement 38
  4.1.1 Immersion Program Data 38
  4.1.2 Adult Natural Corpora Data 39
 4.2 NNS Gender Assignment Data 43
 4.3 NNS Gender Agreement Data 48
Chapter 5: Differences in L1 and L2 Gender Representation and Processing 56
 5.1 Model of Bilingual Language Production 56
 5.2 NS and NNS Gender Congruency Effects 57
 5.3 NS and NNS Reliance on Phonological Cues 58
 5.4 Role of the L1 in L2 Gender Processing 62
 5.5 Summary of L2 Gender Processing 65
 5.6 Grand Summary 67
Chapter 6: Current Study 69
 6.1 Pienemann's Processability Theory and L1-L2 Transfer 71
 6.2 Grammatical Gender Systems 73
  6.2.1 Spanish 74
  6.2.2 Dutch 76
  6.2.3 English 78
 6.3 Research Questions 80
Chapter 7: Methodology 83
 7.1 Participants 83

7.2 Pre-screening Task 84
7.3 Gender Priming Task 89
7.3.1 Materials 89
7.3.2 Design 91
7.3.3 Procedure 93
7.3.4 Pilot 95
7.4 Grammaticality Judgment Task 102
7.4.1 Materials 102
7.4.2 Design 106
7.4.3 Procedure 106
7.4.4 Pilot 110
7.5 Operation Span 116
7.5.1 Materials 118
7.5.2 Design 119
7.5.3 Procedure 119
7.6 Gender Assignment Post-test 120
7.7 Language History Questionnaire 121
7.8 General Procedure 121
Chapter 8: Results 123
8.1 Language History Questionnaire 123
8.1.1 French Native Speakers 123
8.1.2 Spanish Native Speakers 124
8.1.3 Dutch Native Speakers 125
8.1.4 English Native Speakers 127
8.1.5 Comparison of Non-native Speakers 128
8.2 Gender Assignment Post-test 134
8.3 Operation Span 137
8.4 Gender Priming Task 140
8.4.1 French Native Speakers 144
8.4.2 Spanish Native Speakers 156
8.4.3 Dutch Native Speakers 162
8.4.4 English Native Speakers 171
8.4.5 Discussion 176
8.4.6 Conclusion 187
8.5 Grammaticality Judgment Task 188
8.5.1 French Native Speakers 188
8.5.2 Non-native Speakers 196
8.5.2.1 Spanish Native Speakers 197
8.5.2.2 Dutch Native Speakers 202
8.5.2.3 English Native Speakers 209
8.5.3 Discussion 213
8.5.4 Conclusion 224
Chapter 9: General Discussion and Conclusion 226
9.1 Summary of Findings 226
9.1.1 Gender Priming Task 226
9.1.2 Grammaticality Judgment Task 228

9.2 Implications 230
9.3 Additional Considerations.. 234
9.4 Conclusions and Future Directions.. 237
Appendix A.. 239
Appendix B 240
Appendix C 241
Appendix D.. 246
Appendix E 247
Appendix F 253
Appendix G.. 255
Appendix H.. 258
Appendix I 264
Appendix J 266
Appendix K.. 273
Bibliography 295

List of Tables

Table 1 *Phonological Rules in French Gender System (adapted from Surridge, 1993, 1995)* 8
Table 2 *Morphological Rules in French Gender System* 11
Table 3 *Processing Hierarchy of French Gender Agreement* 50
Table 4 *Overview of French, Spanish, Dutch, and English Gender Systems* 79
Table 5 *Gender Priming Task Conditions* 93
Table 6 *Mean RTs (ms) for 5 NS Pilot Participants* 96
Table 7 *Alario et al. (2004, p. 196) Results* 97
Table 8 *Mean RTs (ms) Without Il and Elle Primes* 98
Table 9 *Mean RTs (ms) for 8 NS Pilot Participants* 99
Table 10 *Mean RTs (ms) Without Il and Elle Primes* 100
Table 11 *Grammaticality Judgment Task Example Sentences* 105
Table 12 *Grammaticality Judgment Task Variables* 106
Table 13 *NNS Pilot Participant Grammaticality Judgment Task Accuracy Scores (%)* 112
Table 14 *Grammaticality Judgment Task Filler Accuracy (%) and Range for Low and High Proficiency NNS Pilot Participants* 114
Table 15 *NNS Pilot Participant Accuracy (%) on Cue, No Cue, Close, and Far Grammaticality Judgment Task Target Sentences* 116
Table 16 *Mean Number of Syllables and Mean Word Length for O-Span Words* 118
Table 17 *Language History Questionnaire Data for SNSs, DNSs, and ENSs* 130
Table 18 *NNS Mean Accuracy (% and range) on the Grammaticality Judgment Task, Gender Priming Task, and Gender Assignment Post-test* 132
Table 19 *Math Accuracy (%) and Absolute Span Scores for All Four Language Groups* 138
Table 20 *Number of Participants (n) and Mean (M) Absolute Span Score (Range) for Low and High Span Groups* 139
Table 21 *FNS Final Model: Significant Main Effects* 145
Table 22 *Characteristics of Ambiguous, Unambiguous, and Exception Target Nouns* 152
Table 23 *Distribution of Primes Across Ambiguous, Unambiguous, and Exception Target Nouns* 154
Table 24 *SNS First Analysis Final Model: Significant Main Effects and Interactions* 157
Table 25 *SNS Second Analysis Final Model: Significant Main Effects and Interactions* 159
Table 26 *DNS First Analysis Final Model: Significant Main Effects and Interactions* 163
Table 27 *DNS Second Analysis Final Model: Significant Main Effects and Interactions* 168
Table 28 *ENS First Analysis Final Model: Significant Main Effects and Interactions* 171
Table 29 *ENS Second Analysis Final Model: Significant Main Effects and Interactions* 173
Table 30 *Distribution (%) of Low, Medium, and High Frequency Words for Each NNS Language Group* 179
Table 31 *T-test Results for Gender Assignment (Accuracy and RT) on Ambiguous, Unambiguous, and Exception Target Nouns on Gender Priming Task* 183
Table 32 *Distribution of Participants Who Show Priming Effects Among Group* 185
Table 33 *ANOVA Results for Target Sentence Accuracy for All Four Language Groups* 217
Table E.1 *Mean (Range) for Grammaticality Judgment Task Target Nouns: Word Frequency, Word Length, Number of Syllables* 252
Table H.1 *Experimental Task Check-list* 261
Table I.1 *HLM Model Fit Statistics* 264
Table K.1 *Mean RT (ms) for Lists 1, 2, and 3 for Each Language Group* 275
Table K.2 *Characteristics of Lists 1, 2, and 3* 276
Table K.3 *SNS Language History Questionnaire data and Language Task Performance* 277

Table K.4 *DNS Language History Questionnaire data and Language Task Performance*.........286
Table K.5 *ENS Language History Questionnaire data and Language Task Performance*.........293

List of Figures

Figure 1. Illustration of gender information storage where the lemma is connected to noun class and gender nodes, which in turn, activate appropriate agreement targets, and subsequently, the phonological form. 23
Figure 2. Illustration of the possible lemma-gender relationships at the lemma level; bi-directional links exist between lemmas and gender nodes, and determiner-noun links develop based on frequency 35
Figure 3. Illustration of L2 French gender representation for Spanish, Dutch, and English NSs.. 81
Figure 4. Mean accuracy (%) on the gender assignment task for all four language groups......... 135
Figure 5. Mean RT (ms) on the gender assignment task for all four language groups........ 136
Figure 6. FNS RTs (ms) in congruent, neutral, and incongruent conditions........ 146
Figure 7. FNS RTs (ms) for ambigous, unambiguous, and exception target nouns 151
Figure 8. Mean RTs (ms) for *le/la*, *un/une*, *mon/ma*, and *il/elle* primes 153
Figure 9. SNS Counterbalancing Group by Congruency interaction, with list (1, 2, 3) displayed above each column to indicate which Group saw which list in each Congruency condition 158
Figure 10. SNS interaction between Congruency and Word Ambiguity for participants who have known French for 2-9 years........ 160
Figure 11. SNS interaction between Congruency and Word Ambiguity for participants who have known French for 10-14 years 161
Figure 12. SNS interaction between Congruency and Word Ambiguity for participants who have known French for 15-23 years 161
Figure 13. DNS interaction between Congruency and Word Ambiguity: Group A........ 164
Figure 14. DNS interaction between Congruency and Word Ambiguity: Group B........ 165
Figure 15. DNS interaction between Congruency and Word Ambiguity: Group C........ 165
Figure 16. DNS interaction between Congruency and AO 169
Figure 17. ENS RTs (ms) for compatible, incompatible, and neutral primes........ 172
Figure 18. ENS interaction between Prime Compatibility and Years France........ 175
Figure 19. Accuracy (% correct) on ambiguous, unambiguous, and exception target nouns in gender assignment post-test for all four language groups........ 181
Figure 20. RT (ms) on ambiguous, unambiguous, and exception target nouns in gender assignment post-test for all four language groups 182
Figure 21. FNS accuracy (%) on target, incorrect filler, and correct filler sentences 192
Figure 22. FNS RT (ms) on target, incorrect filler, and correct filler sentences........ 193
Figure 23. FNS RT (ms) on sentences in Cue and Gender conditions 194
Figure 24. FNS accuracy (%) on target sentence conditions........ 196
Figure 25. SNS accuracy (%) on target, incorrect filler, and correct filler sentences 199
Figure 26. SNS RT (ms) on target, incorrect filler, and correct filler sentences........ 199
Figure 27. SNS RT (ms) on sentences in Gender condition for low and high span groups......... 201
Figure 28. SNS accuracy (%) on target sentence conditions........ 202
Figure 29. DNS accuracy (%) on target, incorrect filler, and correct filler sentences........ 204
Figure 30. DNS RT (ms) on target, incorrect filler, and correct filler sentences........ 204
Figure 31. DNS accuracy (%) on target sentence conditions........ 205
Figure 32. DNS RT (ms) on sentences in Cue and Gender conditions 207
Figure 33. DNS RT (ms) on sentences in Cue and Gender conditions for low and high span groups 208
Figure 34. ENS accuracy (%) on target, incorrect filler, and correct filler sentences 210
Figure 35. ENS RT (ms) on target, incorrect filler, and correct filler sentences 211
Figure 36. ENS accuracy (%) on target sentence conditions 212

Figure 37. Accuracy (%) on target, incorrect filler, and correct filler sentences for all four language groups 214
Figure 38. Accuracy (%) on grammaticality judgment target nouns in gender assignment post-test for all four language groups 216
Figure 39. Accuracy (%) on masculine and feminine target nouns in gender assignment post-test for all four language groups 219
Figure 40. SNS, DNS, and ENS accuracy (%) on sentences in the Cue condition 222
Figure K.1. SNS Counterbalancing Group by Congruency interaction, with list (1, 2, 3) displayed above each column to indicate which Group saw which list in each Congruency condition 274
Figure K.2. SNS interaction between Prime Compatibility and Word Ambiguity: low frequency words........279
Figure K.3. SNS interaction between Prime Compatibility and Word Ambiguity: medium frequency words 279
Figure K.4. SNS interaction between Prime Compatibility and Word Ambiguity: high frequency words........280
Figure K.5. SNS interaction between Prime Compatibility and Word Ambiguity for participants who had spent .2-1.5 years in France........281
Figure K.6. SNS interaction between Prime Compatibility and Word Ambiguity for participants who had spent 2- 3.5 years in France........282
Figure K.7. SNS interaction between Prime Compatibility and Word Ambiguity for participants who had spent 4-14 years in France........282
Figure K.8. SNS interaction between Counterbalancing Group and Word Ambiguity........284
Figure K.9. DNS Counterbalancing Group by Congruency interaction, with list (1, 2, 3) displayed above each column to indicate which Group saw which list in each Congruency condition 285
Figure K.10. DNS interaction between Word Ambiguity and Word Frequency: Group A 287
Figure K.11. DNS interaction between Word Ambiguity and Word Frequency: Group B 287
Figure K.12. DNS interaction between Word Ambiguity and Word Frequency: Group C 288
Figure K.13. DNS interaction between Prime Compatibility and Word Ambiguity: low frequency words........289
Figure K.14. DNS interaction between Prime Compatibility and Word Ambiguity: medium frequency words 289
Figure K.15. DNS interaction between Prime Compatibility and Word Ambiguity: high frequency words 290

Chapter 1: Introduction

One of the most difficult challenges of learning French as a foreign language is mastering the gender system. Native speakers (NSs) of French acquire this system in early childhood, usually by the age of three (Karmiloff-Smith, 1979), but non-native speakers (NNSs) are rarely able to achieve native-like use of gender (Carroll 1989; Holmes & Dejean de la Bâtie, 1999; Surridge, 1995; Tucker, Lambert, & Rigault, 1977), even at advanced levels of proficiency (Bartning & Schlyter, 2004; Dewaele & Véronique, 2000, 2001).

Although there are theories that account for how French NSs master their gender system, it is not fully understood why second language (L2) learners are unable to achieve native-like mastery. The current study addresses this difference in ability between French NS and NNSs; however, before presenting the specific research aims of this paper, it is necessary to clarify four key terms that define the tasks faced by learners of French, both native and non-native.

Gender assignment refers to the gender (i.e., masculine, feminine) of a noun. For example, the French noun *table* (table) is feminine and the French noun *livre* (book) is masculine. *Gender agreement* refers to the appropriate gender marking on determiners and adjectives, and in some languages, verbs, numbers, etc. For example, the determiner and adjective are marked as feminine in the French sentence *<u>La</u> table est blan<u>che</u>* (The table is white). *Gender representation* refers to how the gender of a noun is stored. Dewaele and Véronique (2001) define gender as “an idiosyncratic diacritic feature of French nouns, the value of which has to be acquired individually for every lexical entry stored in the mental lexicon” (p. 276). In other words, the gender feature is stored in the same way as other syntactic properties, such as syntactic category (i.e., noun, verb). However, whether NNSs are able to store gender information as an inherent property[1] of the lexical entry is not clear, and was one of the main research questions of this study. *Gender processing* refers to the realization of gender agreement during oral and written speech production and comprehension. Gender assignment and gender agreement are phenomena of the language itself, whereas gender representation and gender processing are phenomena driven by individuals using the language.

NSs of French seem to have access to word-final cues to gender, termed gender-ending regularities here, and they are able to rely on this knowledge when asked to assign gender to new and nonce words (Holmes & Dejean de la Bâtie, 1999). It is thought that children and adult NSs are able to use this knowledge to acquire a noun's gender (Tucker, Lambert, & Rigault, 1977), but once a word, along with its gender feature, is acquired, gender representation is not reliant on knowledge of gender-ending regularities. NNSs are also able to use gender-ending regularities to assign gender to French words (Hardison, 1992; Holmes & Dejean de la Bâtie; Holmes & Segui, 2006), but this ability to assign gender correctly based on explicit knowledge does not indicate native-like gender representation, nor does it account for typical NNS gender agreement accuracy, which falls below that of NSs.

Models of speech production (De Bot, 1992; Levelt, 1989; Schriefers & Jescheniak, 1999) and recent research on first language (L1) gender processing (Bordag, Opitz, & Pechmann, 2006; Dewaele & Véronique, 2000; Hohlfeld, 2006; Holmes & Segui, 2006) indicate that NSs do not rely on gender-ending regularities during gender processing, but

[1] The definition of "inherent property" will be discussed in Chapter 3.

rather activate inherently stored gender information, a finding which could prove to be key in understanding why NNSs are unable to fully master gender agreement. In other words, NNSs may be able to assign gender in French in a similar manner to NSs, but achieving accurate gender agreement appears to require a different type of processing that is apparently not readily available to NNSs. Furthermore, recent studies on gender processing indicate that a close relationship between the gender systems of a learner's L1 and L2 may be advantageous in terms of achieving native-like processing (e.g., Sabourin & Stowe, 2008, who a found a native-like P600 effect during gender processing for German L1 learners of Dutch, but not for Romance L1 learners of Dutch), whereas learners whose L1 gender system is not similar to that of the L2 may compensate by relying on surface cues, such as the gender-ending regularities used during gender assignment (Bordag et al.), and non-linguistic properties, such as memory, to maintain gender information throughout a sentence in order to carry out accurate gender agreement (Holmes & Dejean de la Bâtie, 1999; Sabourin & Stowe, 2004).

However, no studies to date have examined both gender representation and gender processing while taking into account L1-L2 similarity. Therefore, the purpose of this dissertation research was is to investigate how learners of French represent grammatical gender, and how this representation corresponds to processing of grammatical gender. Specifically, this research investigated differences between L1 and L2 gender representation, the relationship between a learner's L1 gender system and their L2 gender representation and processing, and the role of gender cues and non-linguistic processing constraints, such as working memory (WM), during L2 gender processing.

Chapter 2 describes the French gender system, including both gender agreement marking and gender assignment patterns based on phonological, morphological, and semantic rules, and current theories of how NS children use these regularities to acquire L1 gender. Chapter 3 presents a model of L1 language processing specific to grammatical gender, followed by an overview of recent studies on L1 gender processing. Chapter 4 reviews the literature on L2 gender assignment and agreement ability, and Chapter 5 presents the differences in L1 and L2 gender representation and processing. Chapter 6 introduces the current study and research questions, and Chapter 7 presents the experimental tasks. Chapter 8 presents the results, and Chapter 9 presents the general discussion and conclusion.

Understanding how L2 learners of French represent and process grammatical gender will contribute to our knowledge of the role of the critical period and transfer in SLA. If L2 learners are unable to develop native-like representation and/or processing of grammatical gender, it suggests that this component of SLA is age-sensitive and merits further investigation as to when the cut-off for native-like ability occurs, and whether this cut-off is consistent across L1s. In addition, research in this area has pedagogical implications in terms of setting expectations for students of French, developing strategies to overcome non-native-like gender agreement accuracy, and designing appropriate and effective teaching materials.

Chapter 2: French Gender System and L1 Gender Acquisition

2.1 French Gender System

Gender is the division of nouns into classes based on phonological cues and/or semantic properties, such as sex and animacy (O'Grady & Guzman, 2001). The complexity of gender systems varies across languages, with some languages, such as English, employing only a pronominal gender system, and other languages, such as the Bantu language SiSwati, distinguishing among more than a dozen gender classes, which may be marked on adjectives, verbs, adverbs, numerals, and conjunctions (Corbett, 1991). The French gender system assigns masculine and feminine gender to all nouns, and gender is marked on determiners, pronouns, and adjectives. For example, the sentence, *The small table is old*, which is unmarked for gender in English, contains three feminine gender markings in French:

(1) L<u>a</u> petit<u>e</u> table est viei<u>lle</u>. [l<u>a</u> pətit<u>t</u> tabl ɛ vj<u>ɛj</u>].

That is, a feminine noun requires a feminine determiner and feminine adjectives. However, not all French gender markings have distinct masculine and feminine forms. For example, gender is marked in the singular definite and singular indefinite determiners, but is neutralized in the plural definite and indefinite determiners, as shown in examples 2-5 below.

(2) le lit, un lit [masc. sg.]

the bed, a bed

(3) les lits, des lits [masc. pl.]

the beds

(4) la table, une table [fem. sg.],

the table, a table

(5) les tables, des tables [fem. pl.]

the tables

Gender agreement is marked on adjectives, typically by an additional suffix at the end of the feminine adjective. However, the difference between the masculine and feminine forms may not be realized in their oral and/or written forms. Riegel, Pellat and Rioul (1994) identify three categories of adjectives: (1) adjectives that do not differ in either written or oral language (examples 6 and 7), (2) adjectives that differ only in written, but not oral, language (examples 8 and 9), and (3) adjectives that differ in both written and oral language (examples 10 and 11). According to Riegel et al., two thirds of French adjectives do not distinguish between the masculine and feminine in their oral form, and half do not distinguish between the masculine and feminine in their written form. In other words, only one third of adjectives have distinct masculine and feminine forms in spoken French. Finally, in contrast to determiners, adjectives do not neutralize in the plural.

(6) Le livre est difficile [difisil].

The book is difficult.

(7) La situation est difficile [difisil].

The situation is difficult.

(8) Le livre est cher [ʃɛʀ].

The book is expensive.

(9) La voiture est chère [ʃɛʀ].

The car is expensive.

(10) Le lit est petit_ [pəti].

The bed is small.

(11) La table est petite [pətit].

The table is small.

At first glance, whether a given French noun is masculine or feminine appears to be arbitrary. However, within the past thirty years, patterns of gender assignment have been identified and codified in an attempt to better understand how gender is acquired by both NSs and NNSs of French (Surridge, 1993, 1995; Tucker et al., 1977). The following three categories of rules have been posited to account for the French gender system: phonological rules, morphological rules, and semantic rules.

2.1.1 Phonological Rules

Phonological rules depend on the phonological ending of the noun, as summarized in Table 1 (Surridge, 1993, 1995). Clearly, gender patterns based on phonological rules exist; but as the percentages indicate, except for a handful of phonemes, some are far from reliable. For example, words ending with the phoneme [œ̃] are always masculine, whereas only 83% of words ending in [a] are masculine.

Table 1

Phonological Rules in French Gender System (adapted from Surridge, 1993, 1995)

Predominantly masculine endings					
Phoneme	% Masc.	Example	Phoneme	% Masc.	Example
[œ̃]	100%	parfum (perfume)	[f]	89%	tarif (price)
[ɛ̃]	99%	bain (bath)	[u]	88%	bijou (jewel)
[ã]	99%	banc (bench)	[a]	83%	débat (discussion)
[Ø]	97%	pneu (tire)	[ʀ]	75%	cigare (cigar)
[o]	97%	tricot (knitting)	[ɡ]	73%	catalogue (catalogue)
[ʒ]	94%	étage (floor, of a bldg)	[y]	72%	tissu (fabric)
[m]	92%	problème (problem)	[k]	67%	parc (park)
[ɛ]	90%	billet (ticket)	[b]	65%	microbe (germ)

Predominantly feminine endings					
Phoneme	% Fem.	Example	Phoneme	% Fem.	Example
[z]	90%	chose (thing)	[j]	68%	bataille (battle)
[i]	83%	comédie (comedy)	[ʃ]	66%	cloche (bell)
[ɔ̃]	70%	chanson (song)	[d]	62%	étude (study)
[n]	69%	lune (moon)	[s]	62%	chance (luck)
[v]	69%	grève (strike)	[ɲ]	61%	montagne (mountain)

Gender ambiguous endings					
Phoneme	% M/F	Example	Phoneme	% M/F	Example
[p]	≈ 50%	nappe (fem.) (tablecloth) cap (masc.) (cape)	[t]	≈ 50%	doute (masc.) (doubt) patate (fem.) (spud)
[e]	≈ 50%	cahier (masc.) (notebook) journée (fem.) (day)	[l]	≈ 50%	file (fem.) (line) fil (masc.) (thread)

Furthermore, there is a high proportion of exceptions to the phonological ending rules, many of which are high frequency words. Some examples of gender rule exceptions include: *fruit* (masc., fruit), *camion* (masc., work), *page* (fem., page), and *pomme* (fem., apple). Thus, though phonological rules clearly provide gender information in many cases, they are not reliable.

2.1.2 Morphological Rules

Gender assignment based on morphological rules depends on the lexical structure of the word. Unlike the variability noted in gender assignment based on phonological rules, "*les règles morphologiques sont valables à 100% pour les noms formés correctement*" (morphological rules are 100% reliable for correctly formed nouns) (Surridge, 1989, p. 37). Compound nouns, suffixation, and grammatical conversion are all governed by morphological gender rules. Compound nouns, which may be formed by combining two nouns (*presse-citron* [lemon squeezer]), a noun plus a prepositional phrase (*fil-de-fer* [wire]), or a noun plus an adjective (*coffre-fort* [safe]), always take the gender of the first noun (Surridge, 1993). Therefore, the phonological ending of the first noun, rather than the word's phonological ending, provides the gender cue.

Nouns may be created by suffixation, for example, by adding a suffix to a verb, a verb stem, an adjective, a past participle, or an existing noun. Gender rules based on noun suffixation are illustrated in Table 2 (Surridge, 1995).

Table 2

Morphological Rules in French Gender System

Feminine nouns formed by suffixation	Example	Translation
Verb + *-ation, -ition, -ution, -tion/-ssion*	Admirer/Admiration	Admire/Admiration
Adjective + *-té, -ité*	Beau/Beauté	Beautiful/Beauty
Adjective or participle + *-ance, -ence*	Absent/Absence	Absent/Absence
Noun + *-erie, -ie*; verb + *-erie*	Libraire/Librairie	Bookseller/Bookshop
Past participle + *-e* (feminine or silent e)	Allée/Allée	Went/Path
Verb (stem or past participle) + *-ure*	Coiffer/Coiffure	Do sb's hair/Hairstyle
Feminine form of adjective + *-eur, -esse*	Grande/Grandeur	Great/Greatness
Noun + *-ette*	Camion/Camionnette	Truck/Small van
Verb stem + *-sion*	Décider/Décision	Decide/Decision
Noun denoting number + *-aine*	Dix/Dixaine	Ten/Ten or so
Noun + *-ée*	Jour/Journée	Day/Day (duration)
Masculine nouns formed by suffixation	Example	English Translation
Cardinal number + *-ième*	Six/Sixième	Six/Sixth
Verb + *-ment*	Changer/Changement	Change/Change
Verb + age; noun + *-age*	Barrer/Barrage	Block/Barricade
Noun +*-er, -ier*	Pomme/Pommier	Apple/Apple tree
Verb stem + *-oir*	Raser/Rasoir	Shave/Razor
Noun + *-on*	Balle/Ballon	Ball/Ball (diminutive)

Simple grammatical conversion is another source of new nouns, which occurs when a part of speech other than a noun is used as a noun without any overt changes to the word.

These nouns are always masculine regardless of the phonological rules, for example, *le froid* (the cold) (Surridge, 1995). In cases in which the morphological cue contradicts the phonological cue, as in *ballon (m),* which has the feminine ending [ɔ̃], the morphological rule always trumps the phonological rule. This is why gender predictions based on morphological rules are much more reliable than predictions based on phonological cues.

2.1.3 Semantic Rules

Semantic rules are considered to be the most confusing, often due to seemingly arbitrary 'semantic categories' (Surridge, 1993, 1995). Whereas animate nouns referring to humans are obvious in terms of gender (*la fille* [the girl]*, le garçon* [the boy]), inanimate semantic categories are not. Surridge (1989, 1995) cites 17 semantic categories in French as follows: masculine groups include days of the week, months, seasons, points of the compass, languages and dialects, chemical elements and metals, letters of the alphabet, number nouns, metric measures, colors, wines and cheeses, types of aircraft, and trees; feminine groups include feasts and Saints' days, academic disciplines, types of car, and names of fruit derived from names of trees. There are few exceptions to these semantic category rules; Surridge notes that three types of trees, *yeuse* (holm oak), *épinette* (spruce), and *sapinette* (spruce), are feminine; the fruit *abricot* (apricot), derived from *abricotier* (apricot tree) is masculine; and the type of cars *sedan* (sedan) is feminine. However, semantic rules are considered highly reliable, and in cases of semantic-phonological conflict, the semantic rule dominates. For example, *le lundi* (Monday) conflicts with the predominately feminine phonological ending [i] rule. In this case, the phonological rule ([i] is a feminine ending) is violated, but the semantic rule is not. The potential importance of this hierarchy will become clearer in the next section, which considers the acquisition of gender.

To summarize, the French gender assignment system is governed by phonological, morphological, and semantic rules. However, the overall system is not one hundred percent reliable due to the hierarchy of rules (i.e., morphological rules dominate phonological rule) and to the existence of exceptions in some high frequency words (i.e., *la cage* [the cage]). How L1 learners make use of this system will be addressed in the next section.

2.2 L1 Gender Acquisition

The phonological, morphological and semantic patterns outlined above are key in understanding how NSs of French acquire and process gender. Although NSs are often unaware of the phonological and morphological patterns and offer alternative explanations for their gender assignment ability -- for example, having learned each noun along with its appropriate article (Tucker et al., 1977) -- Clarke (1985) and Tucker et al. maintain that NS gender acquisition is indeed based on a rule-governed system. Specifically, they reject the possibility of NSs learning nouns and gender as associated pairs, claiming that this would entail the child having to learn, for example, the word *porte* (door) as *laporte* (the-door), *cetteporte* (this-door), *uneporte* (a-door), etc. This task would not only be impractical and nearly impossible, it would affect the child's use of determiners in sentences. It would also not be able to account for accurate gender agreement throughout the sentence. It appears, then, that despite a NSs' lack of awareness of gender patterns, NS gender acquisition is a processing phenomenon that is based on a rule-governed system.

2.2.1 L1 Gender Acquisition Data

Tucker et al. (1977) were the first researchers to examine how child and adult NSs of French assign gender to nouns. They conducted a series of experiments in an attempt to determine whether NSs rely on noun endings (final phoneme[s]) to assign gender. In these experiments, French NSs between the ages of 7 and 17 either heard or heard and saw real (common and rare) and nonce (but phonotactically possible) French nouns. The participants were instructed to assign masculine or feminine gender to each noun by selecting “un” or “une” as the appropriate indefinite determiner. The nouns’ final phoneme(s) had varying degrees of gender predictability. Some of the endings conflicted in terms of predicting gender depending on whether the final phoneme was considered in isolation or within the context of the preceding phoneme(s). For example, [R] is a predominantly masculine ending, but the ending [yR] is feminine, corresponding to Surridge’s (1993, 1995) phonological and morphological hierarchy.

Regardless of the mode of presentation (aural vs. aural and orthographic), participants’ gender assignments matched the distributional patterns of gender by noun endings that are found in the *Petit Larousse*.[2] The trends were the same regardless of participant age and whether the noun was real, rare, or nonce. For cases in which a noun’s orthographic presentation is typical of one gender (i.e., the orthographic ending *aie* [ɛ] is typically feminine), but the aural presentation ([ɛ]) is typically masculine, participants who both saw and heard the noun relied on the orthographic presentation, thus, determining the noun to be feminine. The participants who only heard the noun relied on the phonologic presentation, thus, determining the noun to be masculine. For

[2] Tucker et al. (1977) used *Petit Larousse* as the corpus to determine gender predictability by noun ending (phonological and morphological); according to the authors, the results are “very similar” (p. 20) to those found by using Le Français fondamental, which is the corpus used by Surridge (1993, 1995).

cases in which the typical gender of a phoneme in isolation conflicted with the typical gender of that same phoneme within the context of the preceding phoneme(s), participants tended to rely on the larger morphological context to determine gender.

Tucker et al. (1977) conclude, "French native speakers consistently assign gender to rarely occurring real nouns, to invented nouns, and to nonsense[3] nouns in accordance with the distributional regularities of the corpus" (p. 57). That is, the results indicate NSs are not relying on memorized information specific to that word, but rather, they can rely on phonological and morphological noun endings. Furthermore, NSs are able to distinguish between phonological and morphological noun endings in order to determine a noun's gender based on the dominant morphological rule.

Based on these results, Tucker et al. (1977) propose that NSs engage in "backward processing", that is, they identify the ending as a probable gender marker, "and then scan backwards into the word until they can determine in which particular subcontext the terminal phone occurs" (p. 62). Their data are consistent with rules outlined by Surridge (1993, 1995), and "backward processing" corresponds to the primary reliance on morphological rules and a secondary reliance on phonological rules. For example, a word with a predominantly (75%) masculine ending such as [R] will be correctly assigned as feminine if the ending [R] is part of the morphological suffix [yR], as in the noun *allure* (speed), which is derived from the verb *aller* (to go). That is, the learner initially identifies the phonological ending [R], but then scans backwards into the word and realizes that the phoneme [R] is part of the feminine morphological ending [yR].

[3] Nonce nouns include both invented and nonsense nouns: invented nouns were created by adding a suffix (either phonological or morphological) to a noun stem with semantic meaning (i.e., *flor)*, as opposed to nonsense nouns, which were created by adding a suffix to a noun stem devoid of semantic meaning (i.e., *fita*). Neither nonce nor invented nouns violated French phonology.

Surridge (1993) develops this theory further by examining the hierarchical relationship between different types of gender rules, which she claims determines the order of acquisition among NSs of French. While Surridge does not provide L1 acquisition data, she theorizes that the frequency of types of nouns (mono-morphemic vs. multi-morphemic) plays an important role in the hierarchy of rule acquisition. According to Gougenheim's (1958) *Dictionnaire du français fondamental*, most French nouns are mono-morphemic, a smaller number are multi-morphemic, and finally, an insignificant proportion of compounds.[4] Based on these results, Surridge concludes that French children first acquire mono-morphemic words, and, therefore, phonological rules, and later acquire multi-morphemic and compound nouns, and the accompanying morphological rules:

> …it seems highly probable that [NSs of French] acquire first a mixture of simple or prefixed nouns to which the phonic rules will apply….These would then be accompanied or shortly followed by suffixed nouns, the gender of which is governed by clear morphological rules (p. 87).

In other words, she claims that NSs acquire gender rules in stages, beginning with phonological rules, and followed by morphological rules, thus, allowing for complete mastery of the gender system and the rules' hierarchical relationships.

L1 acquisition data providing evidence for children's use of noun-ending regularities in gender assignment comes from a study by Karmiloff-Smith (1979), in which child monolingual speakers of French (ages 3-11 years) were asked to assign gender to nonce

[4] The frequency of nouns acquired in early stages of L1 acquisition of French were taken from an examination of Gougenheim's (1958) *Dictionnaire du français fondamental*; Surridge (1993) points out that this corpus is not necessarily representative of early acquired vocabulary, although it is the most suitable until a more accurate corpus becomes available.

French words that had typically masculine endings, typically feminine endings, or gender ambiguous endings.[5] The nonce words were introduced as pictures of imaginary objects with a verbal gender cue in the form of an indefinite determiner that was congruent (i.e., *un coumeau*) or incongruent (i.e., *une coumeau*) with the noun ending, or with an ambiguous gender cue (*deux coumeaux*). Participants were asked questions about the pictures with the elicited response requiring a gender-marked definite determiner.

The congruent indefinite determine + noun trials (i.e., *un coumeau*) did not pose a problem for any of the participants as they produced a definite article that matched the indefinite determiner on 95% of the trials (range 78%-100%). In the incongruent indefinite determiner + noun trials (i.e., *une coumeau*), the younger participants (ages 3-4) were more likely to produce the definite determiner based on the noun's ending than on the indefinite determiner that had been provided in the masculine determiner condition. Specifically, when the indefinite determiner was masculine and the noun ending feminine, the 3-5 year olds produced the feminine definite determiner on 65% (range 56%-81%) of the trials. Interestingly, in the reverse condition (feminine indefinite determiner and masculine noun ending) 3-4 year olds produced a masculine definite determiner on 59% (range 54%-63%) of the trials, but the 5-year-olds produced a masculine definite determiner on only 7% of the trials. The 6-11 year olds, however, showed evidence of relying on the determiner, as they produced a definite determiner that matched the indefinite determiner, regardless of the incongruent noun ending, on 83% (range 68%-96%) of the trials. Therefore, it seems that children begin by relying on a

[5] Karmiloff-Smith (1979) did not distinguish between phonological and morphological endings that predict gender, nor did she specify the proportion of masculine/feminine nouns with the endings selected. However, within the list of experimental nonce words, there were no apparent conflicts of gender prediction between the phonological and morphological endings.

noun's ending to assign gender, but incorporate syntactic cues, such as the determiner, as they get older.

The child NSs' ability to rely on phonological and morphological noun endings in gender assignment was confirmed in the gender ambiguous cue condition. When no gender-marked determiner cue was provided (i.e., *deux coumeaux*), the 3-8 year olds consistently assigned gender based on the noun endings (percentages for all trials not provided in original study). However, the 9-10 year olds consistently assigned masculine to all nouns, regardless of the noun ending, suggesting that they were not relying on noun-ending regularities. The author comments that these participants indicated in a post-task discussion that masculine was similar to a 'neuter' or unmarked gender for unknown nouns with no article cue (p. 160).[6]

Overall, the author concludes that for very young children (ages 3-4 years), gender assignment is based on a noun's phonological or morphological ending, that is, phonological procedures, and that, "it would appear that the phonological procedures are gradually… replaced by… the more foolproof syntactic ones, since consideration of the articles in conflictual situations does indeed increase with age" (p. 167). Together, the findings from Karmiloff-Smith (1979) and Tucker et al. (1977) indicate that during L1 acquisition of French, children are able to rely on the phonological and morphological rules to assign gender to nouns, and for assigning gender to novel nouns, this ability continues into adulthood.

[6] Data on the performance of the 11 year olds on the gender ambiguous cue condition were not reported; therefore, it is not clear from this study whether older children continue to assume masculine gender for novel nouns. However, data from Tucker et al.'s (1977) study indicates this is not the case.

2.2.2 L1 Gender Assignment Data in adults

More recently, adult NSs' ability to assign gender to nonce words was examined in a simple gender assignment task (Holmes & Dejean de la Bâtie, 1999). The NSs, 44 psychology students at a university in France, were presented with 68 real French nouns whose endings were either typically masculine or typically feminine, and 34 "exceptions", which were real French nouns whose endings were atypical for their gender. For example, *peau* (skin) is considered an exception because it has a typically masculine ending, but is a feminine word. In addition, 102 nonce nouns with endings typical of either masculine or feminine endings (81-100% predictability) were included. Both phonological and morphological noun endings were included; however, they were not analyzed separately, and, in some cases, the typical gender of the final phoneme conflicted with the typical gender of the morphological ending of which it was a part. The words appeared for 1000ms in the middle of a computer screen, and participants had 6000ms to indicate by pushing a button whether the word was masculine or feminine. Participants achieved 96.1% accuracy on the regular nouns, 91% accuracy on the exceptions, and 80% accuracy on the nonce nouns (based on the typical gender of the nonce noun's ending). The participants were also slower, in addition to being less accurate, at assigning gender to nonce words.

The authors take the slower reaction times (RTs) and lower accuracy on nonce nouns as evidence that "explicit knowledge of ending rules played little part in the native speakers' gender attributions to real words" (p. 499). In other words, the slower and less accurate gender assignment on nonce nouns, for which NSs had to rely on noun endings, compared to the faster and more accurate gender assignment for real nouns (both regular and exception) suggests NSs were not relying on noun endings to assign gender to real nouns. Furthermore, the NSs were more likely to attribute masculine to nonce nouns, unless the noun ended in the letter “e”; this tendency most likely reflects the assignment of masculine to new words that enter the French language. That is, NSs treat the nonce noun as they would a new word, for example, one borrowed from another language, and determine it to be masculine. The authors suggest that the tendency to rely on the final letter “e” as a marker of feminine may be due to the NSs judging the nonce words as adjectives, rather than nouns. To rule out this possibility, they conducted a similar follow-up task in which participants were presented with determiner + nonce noun phrases and asked to judge the phrases as correct or incorrect. This design ensured that the nonce words were considered as nouns. Participants were able to rely on the noun endings. They achieved 78.2% accuracy on the nonce nouns with either masculine or feminine typical endings (74.7% for feminine, 81.8% for masculine). Overall, the slightly lower accuracy rate on the exception nouns and the high accuracy on the nonce nouns indicate that adult NSs do rely on word ending cues to assign gender, at least for unknown and nonce words.

In sum, the use of phonological, morphological, and semantic rules explains how NSs acquire the gender classification system and apply it to new words. It does not, however, explain the NS’s gender agreement accuracy realized in speech production. Understanding how gender is stored and retrieved for production is a debated issue. Current theories of speech production claim that gender agreement realization is a function of a storage and nodal system at the lexical level. An overview of a gender storage and nodal model, as well as studies designed to test it, will be presented in Chapter 3.

Chapter 3: Model of Grammatical Gender Processing

3.1 Gender Processing within Levelt's Model of Speech Production

Despite gender assignment being governed by systematic rules, it is commonly accepted that a noun's gender feature, once acquired, is stored as an arbitrary and fixed feature of the noun. For example, Schriefers, Jescheniak, and Hantsch (2002) refer to grammatical gender as a feature that is "lexically specified and hardwired" (p. 942); in other words, grammatical gender is an inherent property of the noun, which allows for fast and accurate access of a noun's grammatical gender (Schriefers & Jescheniak, 1999). This view is reflected in one of the most prominent models of language production, Levelt's model of speech production (Levelt, 1989; Levelt, Roelofs, & Meyer, 1999), in which grammatical gender is represented as a lexical-syntactic property of the noun.

Levelt's (1989) model of speech production consists of several processing components in which different aspects of speech production are formulated. Briefly, the preverbal message is formed in the Conceptualizer and sent to the Formulator, which carries out grammatical and phonological encoding of the preverbal message in order to create an internal speech plan. The Articulator automaticallyconverts the speech plan into overt speech.

Because this dissertation research looks at representation and processing of grammatical gender, the processes that take place during grammatical encoding are central to the main proposal. Specifically, grammatical encoding involves creating a surface structure by accessing lemmas that contain the lexical meaning and the syntactic information (e.g., argument structure, such as: "give" is a verb which can take a direct object, indirect object, etc.). This information is then submitted to phonological encoding, which is a separate process that takes place after grammatical encoding. Within this model, gender agreement is processed through grammatical encoding.

While the definition of grammatical gender as an inherent property or fixed feature of a noun implies that gender is an inseparable characteristic of a lemma, according to Levelt's model (1989; Levelt, Roelofs, & Meyer, 1999), rather than specifying grammatical gender individually for each noun, all nouns are linked to the appropriate abstract gender node. In this model, illustrated in Figure 1,[7] the lemma node (e.g., the lexical item) is connected to a syntactic node, which encodes the word's syntactic category (e.g., noun, adjective, verb), and to a gender node, which encodes the word's grammatical gender. All nouns of the same grammatical gender share a connection to that gender node, which is connected to the appropriate agreement targets for that gender. Separately, each lemma is also connected to phonological nodes that encode the phonological form, and which are activated after all syntactic information is made available. Therefore, when a lemma is selected, its grammatical gender is activated via a nodal link prior to and independently from the phonological form.

[7] This model is a simplified version adapted from Schriefers & Jescheniak (1999).

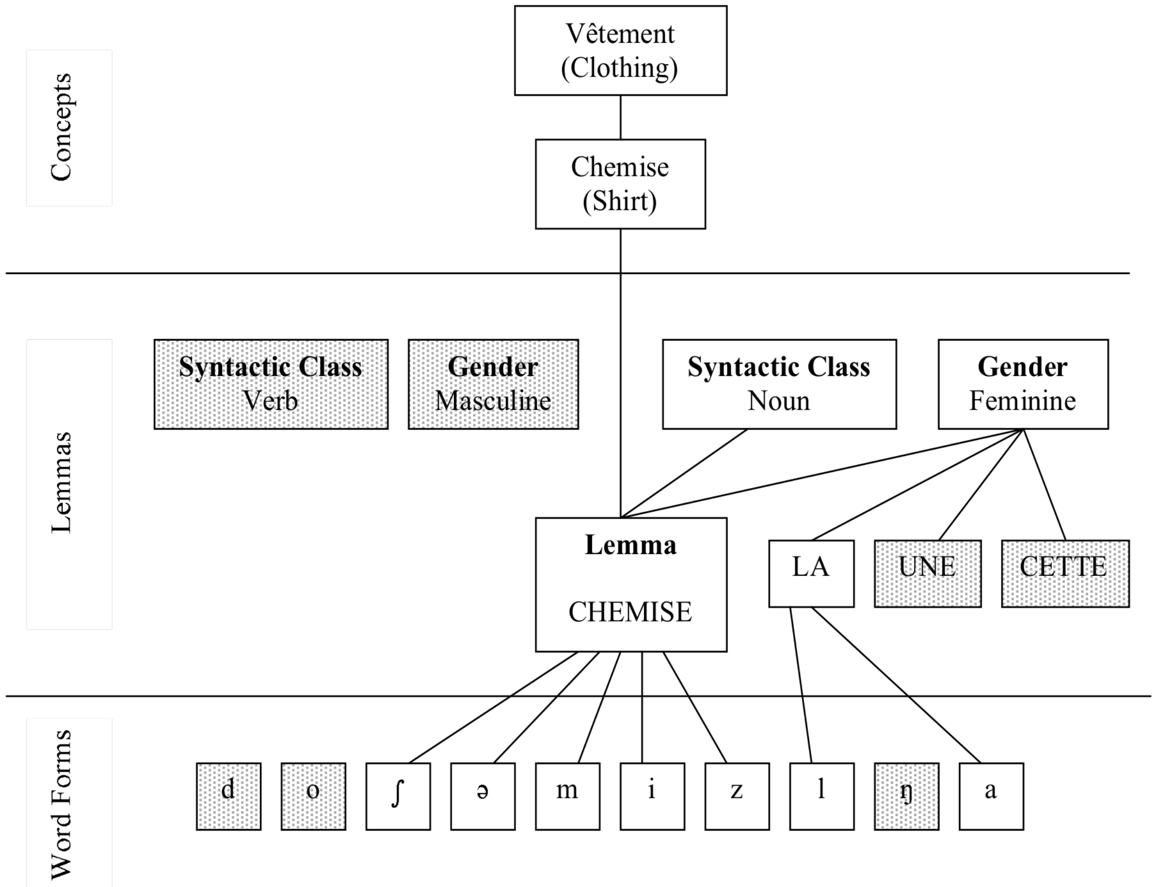


Figure 1. Illustration of gender information storage where the lemma is connected to noun class and gender nodes, which in turn, activate appropriate agreement targets, and subsequently, the phonological form.

The following section reviews studies on L1 grammatical gender processing that attempt to test this model. While results support the general architecture of a gender-nodal system, the exact relationship between the lemma and its gender information remains ambiguous. The implications of this ambiguity for the current study will be addressed at the end of the section.

3.2 L1 Gender Processing Studies

3.2.1 Empirical support for Levelt's model: Physiological studies

One way to determine whether grammatical gender is processed as a syntactic phenomenon is to examine brain responses to gender violations. An event-related potential (ERP) is an electrophysiological response to a stimulus. One type of ERP waveform is the P600, which is a positive deflection in the brainwave that reaches its peak around 600ms after the presentation of an ungrammatical word. P600 effects have been found for a range of syntactic violations, including subcategorization and phrase structure violations, violations of subjacency, and most relevant to this dissertation, number, gender, and case violations (Hagoort, Brown, & Osterhout, 1999). Hagoort and Brown (1999) looked at ERPs for an online Dutch grammaticality judgment task containing determiner-noun gender agreement errors. Dutch NSs saw the sentences presented one word at a time in the middle of a screen and indicated by pushing a button whether the sentence was grammatical or ungrammatical. Each ungrammatical sentence had a grammatical counterpart to which ERP responses were compared. Participants achieved 97% accuracy on the ungrammatical sentences containing gender agreement errors (no other errors existed in the sentences), indicating that they were sensitive to these errors. In addition, a clear P600 response was found for these sentences, but not for their grammatically correct counterparts, providing evidence that establishing determiner-noun gender agreement is a syntactic process. The authors conclude that,

> …the processing of [a noun's] lexically frozen gender features, and more particularly the unification of the gender feature value of the noun with

that of the definite article during on-line processing of noun phrases, seems to be a syntactic process (p. 725).

Within Levelt's model presented in Figure 1, determiner-noun gender agreement is a syntactic process that occurs at the lemma level.

Another key point of the model is the separate activation of a lemma's grammatical and phonological information. Findings of several studies on gender processing are consistent with the claim that information about a word's gender is activated before and independently of the word's phonological form. That is, processing grammatical gender takes place at the lemma level and is not dependent on the word's phonological form. For example, Vigliocco, Antonini and Garett (1997) found that in a tip-of-the-tongue (TOT) state, NSs of Italian were able to identify the gender of the target noun when they were unable to produce the word.[8] That is, they had access to the grammatical gender even when they did not have access to the phonological information.

Physiological evidence has shown that grammatical gender becomes available before phonological information. Van Turennout, Hagoort and Brown (1999) conducted a go/no-go picture naming task and recorded lateralized readiness potentials (LRPs) of hand motor responses, that is, brain activity that occurs during the preparation of a physical movement. In one version of the task, Dutch NSs saw colored line drawings and were told to determine the picture's gender by pushing one of two buttons ("de" for common gender or "het" for neuter gender), only if the noun started with a specific phoneme (go-trial), for example, if the noun began with a /b/. If the noun did not begin

[8] Participants were given a definition and asked to produce the corresponding word. When a participant was in a TOT state (i.e., knew the word but was unable to retrieve it), he/she was asked to indicate on paper his/her "guess" of the word's gender. After making a guess, the participant was presented with the word and asked to confirm that this was the word he/she had been thinking of.

with a /b/, the response had to be withheld (no-go trial) and no button pushed. In a second version, the participants were told to determine the picture's initial phoneme by pushing one of two buttons ("b" or "s"), but only if the noun had a specific gender (go-trial), for example, if the noun was a "het" (neuter gender) noun. If the noun was a "de" (common gender) noun, the response had to be withheld (no-go trial) and no button pushed. LRPs developed in the first version for both the go and no-go trials, demonstrating that the participant was preparing to determine the noun's gender, even when the initial phoneme indicated a no-go trial. In the second version, however, LRPs developed only in the go-trials, demonstrating that the participant only prepared to determine the initial phoneme when the noun's gender indicated a go trial. Thus, in the first version, information about the word's gender was transmitted and a motor response prepared, and subsequently the phonological information was transmitted and either permitted or prevented the motor response from taking place. In the second condition, information about the word's gender was transmitted, and only if the condition permitted, a motor response based on the word's phonological information was prepared and executed. According to the authors, "[t]he findings support the claim that syntactic information influences response preparation at an earlier moment in time than phonological information" (p. 669). By syntactic information, the authors are referring to grammatical gender as a syntactic feature of the word; in other words, a word's gender is available before the phonological form.

3.2.2 Empirical support for Levelt's model: Natural speech production

Errors that occur in natural speech production, rather then being elicited under experimental conditions, also provide evidence that grammatical gender is activated prior to phonological form. Analyses of speech error corpora reveal that in slips of the tongue, in which one word is erroneously substituted for an intended word, syntactic class (i.e., noun, verb), as well as grammatical gender are often preserved. Arnaud (1999) found that out of 449 noun substitutions, 338 (86%) of the intended and substituted words shared the same gender. For example, *frigo* and *lave-vaiselle* are both masculine.

(12) ta vaiselle rouge, elle est pas vraiment faite pour aller au *frigo* (instead of *lave-vaisselle)*

your red dishes, they weren't really made to go into the *fridge* (instead of *dishwasher*) (p. 272)

Although a tendency for substitutions of the same grammatical gender could potentially be explained by a checking mechanism that prevents agreement violations between the substitution and other parts of the sentence that have already been selected and produced (i.e., *au* in the sentence above, which is marked for masculine), Marx (1999) found that grammatical gender is preserved in noun substitutions even when the context does not require a gender-marked word. Substitution errors in which grammatical gender is preserved suggests that gender information is available before phonological information. Furthermore, it indicates a bi-directional link between the gender node and lemmas. For example, in a semantic substitution error scenario, the intended lemma activates the feminine gender node as well as a cohort of semantically related lemmas. Instead of the intended lemma reaching activation levels that enable subsequent activation of its phonological representation, one of the alternate lemmas is selected instead. However, the activated feminine gender node limits the semantic substitution to other feminine lemmas, and, thus, the grammatical gender is preserved.

3.2.3 Empirical support for Levelt's model: Behavioral studies

Gender distraction and gender priming tasks are used to investigate the relationship between the lemma and its corresponding gender node in a controlled environment. Although facilitation and inhibition effects found in both types of tasks provide evidence for a gender-nodal system during gender processing, both have limitations and are often context sensitive. In gender distraction tasks, a participant is presented with a distractor noun, followed by a target picture to be named that is either gender congruent or gender incongruent to the distractor. Faster naming times in the congruent condition indicate facilitation due to the distractor word and target picture activating the same gender node and, thus, facilitating production of the target picture. Slower naming times in the incongruent condition are a result of gender node competition, in which the distractor word activates one gender node and the target picture activates another, thereby causing competition that slows production of the target picture's name. A series of gender distraction studies provide evidence for the storage and nodal model; however, results are sensitive to both the stimulus onset asynchrony (SOA) and the response required by the participant (i.e., determiner + noun vs. bare noun).

Schriefers (1993) conducted a gender distraction study with Dutch NSs examining gender congruency effects during the production of noun phrases (NPs). Distractor words were presented before (SOA -200ms), simultaneously with (SOA 0ms), and after (SOA +450ms) colored target pictures were presented, and were either gender-congruent or gender-incongruent with the target picture. The participants named the target pictures in a determiner-adjective-noun condition (det. condition), with the determiner marking the noun's gender (e.g., *het groene bed* [the green bed])[9], or in an adjective-noun condition (adj. condition), with the adjective marking the noun's gender (e.g., *groen bed* [green bed]*)*. Schriefers found gender distraction effects (increased time to produce the target picture NP in the gender-incongruent condition) for both the det. and adj. conditions at SOA = 0ms, although the effect for the adj. condition was half the size of the effect for the det. condition. There was an effect for the det. condition when the SOA was -200ms, but no effect for the adj. condition at this SOA, and there were no effects for either condition when the SOA was +450ms. Schriefers explains the difference in effect between the det. and adj. conditions as a result of the adjective stem being selected before its gender-marked suffix, allowing production of the adjective to begin before the gender processing is complete; in the det. condition, production of the determiner cannot begin until gender processing is complete. For example, the first three phonemes of the adjective *groene* (green) may be selected before the final phoneme that carries gender information is determined. Because the determiners *het* and *de* begin with different phonemes, production cannot begin until the appropriate determiner has been selected. On the whole, Schriefers takes the gender distraction effect as evidence for competition between gender nodes activated by the distractor word and target picture.

However, Schriefers' (1993) study only considered conditions in which production of a gender-marked word (either determiner or adjective) was required along with the target noun. It is unclear from this study whether the gender node plays a role in production when gender information is not required in the response. This distinction was addressed by La Heij, Mak, Sander, and Willeboordse (1998), who examined the conditions under

[9] The adjective neutralizes when the determiner is definite, therefore, the adjective is neutral as to the noun's gender.

which gender congruency effects can be obtained. Following the same task design used by Schriefers, they found congruency effects in tasks that required participants to produce the article + noun, but not in tasks that required participants to produce only the bare noun. They infer from these results that either the distractor noun's gender is not activated in tasks that do not require production of an article, or the gender of the distractor noun is activated but does not affect the retrieval of a different, single noun. La Heij et al. also found that familiarity has an effect on the retrieval of the word's gender, which, in turn, affects the effect size of gender congruency. That is, the more familiar the distractor word, the greater the effect size, a finding that "can be accounted for by assuming repetition-sensitive links between lemmas and the gender nodes" (p. 217).

Despite the robust results in Dutch found by Schriefers (1993) and La Heij et al. (1998), the findings have not been replicated for Romance languages (Alario & Caramazza, 2002; Miozzo, Costa, & Caramazza, 2002). The lack of gender congruency effects has been explained by Caramazza, Miozzo, Costa, Schiller, and Alario (2001) as a result of the different information required for determiner selection in Romance languages. Whereas selection of the appropriate determiner in Dutch only depends upon semantic and syntactic information (number, definiteness, and gender), selection of the appropriate determiner in many Romance languages (i.e., Italian, Spanish, French) relies on semantic, syntactic and also phonological information. For example, in French a vowel-initial noun, either masculine or feminine, takes *l'* as the definite article, whereas a consonant-initial masculine noun takes *le*. Therefore, selection of a Dutch determiner occurs at an earlier stage than it does for a determiner in a Romance language, which can only occur after the target noun's phonology has been activated. Consequently, gender congruency facilitation and inhibition will not be evident in Romance languages, as they were in Dutch. However, even with positive SOA times to account for the later time course of determiner selection in Romance languages (as opposed to the negative SOAs that yielded gender congruency effects in Dutch), no gender congruency effects were found for Spanish or Italian (Miozzo et al.).

That gender distraction effects seem to be dependent on the time course of determiner selection suggests that gender node competition may not be driving the results found in Schriefers (1993) and La Heij et al. (1998). A potential alternative explanation was raised by Miozzo and Caramazza (1999), who suggested that determiner competition rather than gender competition could account for the findings. To investigate this possibility, Schriefers, Jescheniak, and Hantsch (2002) designed an experiment, using German, to test whether determiners compete for selection in an object-naming task. German has three gender classes, masculine, feminine, and neuter, each of which has a corresponding singular definite determiner, *der*, *die*, and *das*, respectively. However, plural determiners neutralize into one form, *die*, for all three gender classes. Based on this structure, the authors propose three hypotheses: the number-dominance hypothesis, the gender-dominance hypothesis, and the singular-as-default hypothesis. According to the number-dominance hypothesis, the number feature dominates the gender feature such that a singular noun will activate all three singular determiners, whereas a plural noun will only activate the one plural determiner. The gender feature will then select the appropriate determiner. This hypothesis predicts that selecting a plural determiner, which has only one candidate, and, therefore, only one step, will be faster than selecting a singular determiner, which has three candidates, and, therefore, requires a second step of using the gender feature to select the appropriate candidate.

In contrast, the gender-dominance hypothesis assumes the gender feature dominates the number feature. A masculine target noun will first activate the masculine singular and masculine plural determiners, *der* and *die*, and the number feature will then select the appropriate determiner. This hypothesis makes different predictions depending on the gender of the target noun. Masculine and neuter nouns will activate two determiner forms, the singular (*der, das*) and the plural (*die*), whereas a feminine noun will only activate one form, (*die*), for both singular and plural. Therefore, determiner competition will occur for masculine and neuter nouns, but not for feminine nouns, resulting in faster naming times for feminine nouns than masculine and neuter nouns.

Finally, the singular-as-default hypothesis assumes the singular is the default value such that a picture will activate both the gender and the singular determiner, and a plural feature will then activate the appropriate plural determiner. Therefore, singular masculine and neuter nouns will activate *der* and *das*, and plural masculine and neuter nouns will also activate *die*, whereas singular and plural feminine nouns will only activate *die*. According to this hypothesis, for masculine and neuter nouns, selection of plural determiners will take longer than selection of singular determiners, but for feminine nouns, there will be no difference between singular and plural.

The results from a picture naming task in which participants named either one or two objects with the appropriate definite determiner support the singular-as-default hypothesis. That is, production of plural determiner-noun phrases as compared to singular determiner-noun phrases was slower for masculine and neuter target nouns, but not for feminine target nouns. The authors conclude that determiner competition may account for at least some of the gender congruency effects found in the gender distraction studies.

One way to eliminate the potential confounding role of determiner selection in gender congruency tasks is to eliminate the use of determiners in production. Although congruency effects disappear in bare noun conditions in gender distraction tasks (La Heij, 1998), they are, nevertheless, robust in gender priming tasks, which do not require production of a determiner. Alario, Matos, and Segui (2004) examined gender priming effects in NSs of French. Participants saw a gender congruent or incongruent prime in the form of a definite determiner (*le* or *la*), an indefinite determiner (*un* or *une*), or a possessive adjective (*mon* or *ma*), or a gender-neutral linguistic baseline, *chaque* (each), followed by a target picture to be named. For example, in the congruent condition, a participant would see the prime *le* followed by a target picture of a book (*livre*, masc.). The participants were told to silently read the prime word and name the target as quickly as possible. In addition to the congruent, incongruent, and neutral conditions (48 in total), there were five catch trials in which participants saw a prime followed by a question mark and were asked to repeat the word they had just seem. The purpose of the catch trials was to ensure that the participants were processing the primes throughout the experiment.

The results showed both congruency and incongruency effects. That is, picture naming latencies were shorter in the congruent condition as compared to the neutral condition, and longer in the incongruent condition as compared to the neutral condition. The authors interpret their finding to mean, "…the processing of the prime activated a given gender and this gender in turn has an influence on the process of lexical selection"

(p. 199). However, one limitation to this study is the inclusion of only syntactically compatible prime-target combinations. Because determiner primes are syntactically compatible with the target pictures, meaning the prime-target pairs form syntactically appropriate noun phrases, as in *le livre* (the book), Alario et al. (2004) proposed that at least some of the facilitation effects found in their study may be due to determiner-noun pair co-occurrence frequency effects. However, although it has not yet been tested, Alario et al. predict that, according to the gender-nodal model, a syntactically inappropriate prime, such as *il* (he) or *elle* (she), will still activate the gender node and result in facilitation/inhibition effects.

3.3 Relationship Between Lemma and Gender Information

To summarize the findings thus far, both gender distraction and gender priming tasks show gender congruency effects, indicating that activation of a distraction word or prime activates a gender node, which in turn, either facilitates or inhibits activation, and, subsequently, production of a target noun. These results provide support for a gender-nodal system. However, as mentioned at the beginning of this chapter, the exact relationship between the lemma and its gender information remains unclear.

Taking into account determiner competition and prime-target co-occurrence frequency as potentially confounding variables, it is possible to suppose additional links within the model. That is, determiner competition and prime-target co-occurrence effects do not rule out lemma-gender node links. Figure 2 illustrates the possible relationships between a lemma and its gender information at the lemma level.

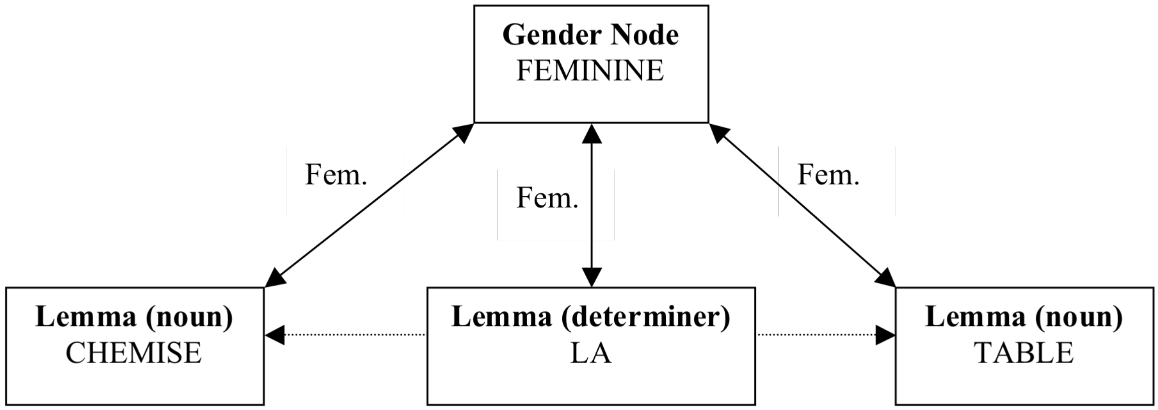


Figure 2. Illustration of the possible lemma-gender relationships at the lemma level; bi-directional links exist between lemmas and gender nodes, and determiner-noun links develop based on frequency

Consistent with the model presented in Figure 1, lemmas are linked to the appropriate gender node such that activation of a lemma automatically activates its gender information. The bi-directional link between the gender node and lemma allows for the lemma to activate the gender node, and vice versa, resulting in gender congruency effects in both gender distraction and gender priming tasks. However, in addition to the lemma-gender and gender-lemma links, determiner-noun links develop and strengthen as a result of frequency, which may also play a role in gender and lemma activation. For example, activating *la* in the form of a prime may activate all frequent feminine nouns via the determiner-noun link, that is, nouns that co-occur with the determiner *la*, in addition to activating the feminine gender node that will also activate feminine nouns. Therefore, the determiner-noun link may provide additional activation of the target noun. However, because *la* only co-occurs with consonant-initial feminine nouns, the additional activation may not take place until the lemma's initial phoneme has been selected at the phonological level; thus, activation as a result of determiner-noun links will occur later in the time-course than activation as a result of determiner-gender node-noun links. Furthermore, whereas the congruency effects found in Schriefers (1993) and La Heij (1998) may be due in part to determiner competition, this finding does not threaten the gender-nodal system. As determiners are lemmas with links to gender nodes, determiner production, like noun production, will be subject to gender congruency effects. In other words, even though determiner competition was reported as a potentially confounding variable in gender distraction tasks, gender congruency may have played a role in production of the determiner, if not the noun.[10]

Finally, there still remains some ambiguity as to how gender information is stored. Dewaele and Véronique (2001) describe gender as a "diacritic feature… which has to be acquired individually for every lexical entry" (p. 276), and Schriefers and Jescheniak (1999) refer to gender as "an inherent property of nouns" (p. 577). Yet the gender information storage system presented in this dissertation proposes that, "instead of specifying each noun's grammatical gender separately in the corresponding lexical entries, there is only one abstract node for each grammatical gender" (Schriefers & Jescheniak, p. 577). These seemingly contradictory representations are not necessarily problematic if one considers the actual link as the inherently stored information. That is, in the same way a lemma has a specification for the syntactic category node (i.e., noun, adjective, verb) to which it is linked, a lemma has a specification for the gender node to which it is linked. Furthermore, the connection between the lemma and the gender node exists within the lemma level, as opposed to the conceptual or phonological level.

[10] Schriefers, Jescheniak, and Hantsch (2002) did not include gender distractors in their task, and, therefore, did not investigate the role of gender congruency effects in determiner selection.

Therefore, regardless of location of the gender information that allows for lemma-gender node connection (i.e., a specification within the lemma itself vs. the link to the node), it is represented at an abstract level of syntactic information, which is the essential distinction for the current study. From here on, the term “inherent” will refer to grammatical gender information that is represented at the lemma level within a gender storage and nodal system, as depicted in Figure 1.

To conclude this chapter, it appears that NSs rely on phonological and morphological regularities during child L1 acquisition (Karmiloff-Smith, 1979) to acquire the gender classification system, which is applied to gender assignment of new words throughout child- and adulthood (Tucker et al., 1977). However, these rules do not appear to play a direct role in gender agreement at the production level; accurate gender agreement during speech production is carried out through a storage and nodal system, which works independently of the phonological form of the word, and, thus, independently of word-ending gender regularities. How the L1 acquisition and processing theories apply to NNSs is not fully understood, although the lack of ability to master the gender system fully, even for highly proficient NNSs of French, suggests that the system is somehow different.

Chapters 4 and 5 review the literature on L2 gender assignment and agreement ability and differences in L1 and L2 gender representation, which in turn, affect gender processing. De Bot’s (1992) model of bilingual language production is offered as an explanation for these differences.

Chapter 4: L2 Gender Assignment and Agreement Ability

4.1 Evidence for Non-native-like L2 Gender Agreement

It is commonly accepted that even highly proficient NNSs are unable to achieve native-like proficiency of the French gender system. Evidence of this deficit comes from English-speaking children in French immersion programs as well as adults who have reached an advanced level of proficiency.

4.1.1 Immersion Program Data

Lapkin and Swain (1977) administered a cloze passage to 173 fifth grade students in Canada: 94 from an early French immersion program, 53 from a Francophone school located in a French-English bilingual community, and 26 from a Francophone school in a monolingual French community. In the early immersion program (for English-speaking children), kindergarten and first grade are taught entirely in French by a French NS; in grades two through five, the curriculum continues to be taught in French, however, students receive 1-2 hours per day of instruction in English (language arts and/or mathematics). The Francophone schools are taught entirely in French by French NSs. The participants completed a 292 word cloze passage in French and filled in 31 blanks, five of which required a French determiner. The overall scores indicate that the early immersion participants (mean score 17.4/31) and the French-English bilingual participants (mean score 18.3/31) performed similarly, whereas the monolingual French participants (mean score 22.1/31) achieved higher scores. These scores are highly correlated (.97) to the French achievement test administered to the early immersion and bilingual participants as part of the schools' requirements. However, the early immersion participants performed worse on the determiner category, providing the incorrect determiner 22.7% of the time, compared to 14.1% and 6.4% by the participants from the Francophone schools in the bilingual and monolingual communities, respectively. According to the authors, these errors were primarily due to incorrect gender agreement. Despite French proficiency similar to French-English bilingual peers, the early immersion participants performed significantly worse on determiner-noun gender agreement.

Naturalistic data from grade-five English NS students in an early French immersion program (in an English speaking community) reveal a similar pattern of gender agreement errors (Harley, 1979). Five students' interview data were transcribed and analyzed and compared with interview data from seven grade-five French-speaking children attending a French school (in a French speaking community). The English NSs achieved 79.1% accuracy (488/617 tokens) on gender agreement, whereas the French NSs achieved 99.4% accuracy (2045/2057 tokens). Furthermore, the English NSs made more errors on adjective agreement than on article agreement, a trend that is found in adult learners of French as well (Bartning 2000; Bartning & Schlyter, 2004; Dewaele & Véronique, 2000, 2001).

4.1.2 Adult Natural Corpora Data

Data from natural corpora show that advanced adult NNSs of French are also unable to achieve native-like gender agreement accuracy. In an interview study (Dewaele & Véronique, 2001), 27 Dutch-French-English trilinguals who had been enrolled in French courses for 4-6 years, and were currently enrolled in a university-level intensive French class, participated in both formal and informal interviews with the researcher. Based on the combined interview data, the participants achieved 94.75 % accuracy on determiner agreement and 91% accuracy on adjective agreement. However, because French adjectives do not always distinguish between the masculine and feminine forms in oral production, it is likely that not all adjective agreement errors were identified, and the 91% accuracy is an over-estimation of their ability.

Furthermore, looking at individual errors, the authors attempted to determine whether adjective agreement errors were due to incorrect gender assignment or incorrect gender agreement. Incorrect gender assignment indicates the participant has incorrectly assigned gender to a word and adjective agreement is based on this incorrect assignment. For example, the authors assume that a participant who utters the phrase "*un affaire religieux"* (a religious affair) has incorrectly assigned masculine to *affaire*, as indicated by the incorrect determiner. Therefore, in this case, the incorrect adjective is a result of incorrect assignment rather than an inability to carry out adjective agreement. On the other hand, a participant who utters the phrase "*une affaire religieux"* has produced the correct determiner, but does not carry out gender agreement on the adjective. Looking at the data for each individual, if a lexical item was used with the wrong determiner, both the determiner and adjective agreement errors were considered to be due to a gender assignment problem. If the lexical item was used with a correct determiner but an incorrect adjective, the adjective error was considered to be due to an agreement problem. Based on these criteria, the authors conclude that approximately 56.6% of the gender agreement errors were due to assignment problems, 13.3% to temporary assignment problems (meaning correct assignment of the same lexical item also occurred in that participant's data), and 23% of the gender agreement errors were due to agreement problems. The remaining 7% of agreement errors fell into the categories of immediate gender error correction, TOT states in which the participant apparently decided to use a different lexical item after the determiner for the original lexical item had already been produced, and transfer of gender from the L1 on high frequency cognates or homophones. Overall, these results indicate that advanced learners do not achieve native-like gender agreement accuracy for determiners or adjectives, and the two main underlying sources are difficulty in assigning gender to lexical items and carrying out gender agreement. Although the authors did not administer a simple gender assignment task (in which participants indicate the gender of isolated French nouns), it would be relevant to determine whether the assignment difficulties existed in an offline task, or if they are a function of online processing of oral production.

Oral interview data from six Swedish learners of French who had studied French for 4.5-6 years and enrolled in university-level French classes also confirm that both determiner and adjective agreement are not native-like (Bartning 2000, Bartning & Schlyter, 2004). These learners achieved 93% and 83% accuracy on definite and indefinite determiner accuracy, respectively, and 81% on adjective agreement. However, these interviews were collected over a two-year period and the participants' gender agreement accuracy during the final interview, when participants were most likely at their highest level of proficiency, was not reported separately. Therefore, it may be the case that these percentages do not represent the participants' true ability.

It is important to note that the proficiency level of the participants in these studies is not clear. The authors describe the participants as "advanced", but no proficiency measure is included. However, Bartning and Schlyter (2004) have also collected interview data from university students who are in training to become teachers of French and students completing doctorates in French, and the French proficiency level of these participants is more likely to qualify as advanced. The authors have not completed the gender agreement analysis for these two groups, but they predict adjective-gender agreement accuracy to be approximately 85%.

Naturalistic data from English L1 near-native learners of Spanish (Franceschina, 2005) show similar results to those of Bartning (2000) and Bartning and Schlyter (2004). Five English L1 adult learners of Spanish who performed within NS range on the University of Wisconsin Spanish Placement Test participated in an informal interview. The learners' determiner-gender accuracy was 92.71% and adjective gender accuracy was 90.35%, as compared to the NS controls' 100% accuracy on both determiner and adjective agreement. The gender agreement errors were analyzed according to the source noun endings, that is, the noun for which the determiner or adjective did not agree. There was no difference in gender agreement accuracy for nouns with gender-typical endings (i.e., feminine nouns ending in the letter "a" and masculine nouns ending in the letter "o") as compared with nouns with ambiguous endings, such as "e", suggesting that the difficulty did not lie in gender assignment, but rather gender agreement. That is, if gender assignment were the source of gender agreement inaccuracy, the results would have shown lower accuracy for source nouns with gender ambiguous endings. Given the similarity of gender systems between French and Spanish (which will be described in detail in Chapter 6), these results provide support for the claim that even highly proficient NNSs have some difficulty with L2 gender agreement.

4.2 NNS Gender Assignment Data

Because L2 learners do not achieve native-like mastery of the French gender system, one might hypothesize that they neither acquire nor process gender in the same manner as NSs. Studies that examine how NNSs assign gender to French words (real and nonce) show that L2 learners use similar strategies to those of French NSs; however, their accuracy still falls behind that of NSs. Specifically, NNSs are able to rely on phonological cues to assign gender. Marinova-Todd (1994, cited in Bialystok, 1997) had English and German L2 learners of French assign gender to French nouns in three gender assignment tasks. The first two tasks involved assigning gender to French real and nonce words based on phonological rules alone; both L1 groups performed equally well and were able to rely on these rules to accurately assign gender. On the third task, a picture containing natural (i.e., semantic) gender information that conflicted with the word's phonological cue was presented along with the word. The English NSs continued to rely only on phonological cues, but the German NSs also incorporated semantic information.[11] That is, the English speakers were similar to the young French children cited in Karmiloff-Smith (1979) in that they relied solely on the noun ending gender cues, whereas the German speakers behaved more like the older children in that they were able to incorporate both noun ending and natural gender cues. These results suggest that both the English and German NNSs of French were able to use phonological cues, but NNSs whose L1 has a gender system similar to that of French were more native-like in that they also relied on natural gender cues when assigning gender in L2.

---

[11] Accuracy scores were not provided in the article, and the author did not provide actual numbers in a request made by personal communication.

Similar evidence of learners relying on noun ending cues has been demonstrated in English learners of Italian. Oliphant (1998) investigated the use of morpho-phonological and determiner cues to determine gender in L2 learners of Italian, which has a similar gender system to that of French. First and second year students of Italian were asked to assign gender to Italian nouns that they were not likely to know. In the first task, nouns were presented aurally and had phonological endings that were typical of either masculine or feminine gender, endings that were ambiguous and did not predict the noun's gender, or morphological endings in which the morphological ending was either congruent or incongruent with the final phoneme. Participants were asked to indicate the gender of each item on an answer sheet. In the second task, nouns were presented aurally with a determiner in five conditions: the determiner was congruent with the noun ending; the determiner was incongruent with the noun ending; the determiner did not contain gender information; the determiner contained gender information, but the noun ending did not; or neither the determiner nor the noun ending contained gender information. Again, participants were asked to indicate the gender of each item on an answer sheet.

Participants were able to rely on phonological cues to assign gender in task one; the mean percent correct for nouns that contain a gender-marked ending was 65.5% (range 19.9% - 99.6%). As evident in the range of accuracy, this ability decreased with the frequency of the ending. For example, (-o) and (-a) are the most common noun endings, and the participants achieved 98.7% accuracy on nouns with these endings. The endings (-i) and (-u) are highly reliable markers of feminine; however, they represent a small group of nouns, and participants achieved only 24% accuracy on nouns with these endings. However, the ending (-i) is also the plural marker for masculine nouns that end in (-o) in the singular form. Therefore, the nouns ending in (-i) may have been perceived as masculine plural nouns rather than feminine singular nouns. Participants achieved 50% accuracy on the nouns with an ambiguous ending. Participants performed poorly on nouns with morphological endings that conflicted with the typical gender associated with the final phoneme. For example, the morphological endings (-ma), (-ista), and (-cida) are all masculine despite the typically feminine final phoneme (-a). Participants achieved 8% (range 1.6%-13.3%) accuracy on these nouns. However, it is important to keep in mind that the participants in this study were beginners and may not have had enough exposure to the gender regularities of morphological and phonological noun endings.

In the second task, accuracy was highest when the determiner was congruent with the noun ending (99.2%). Participants also achieved high accuracy when either the determiner or the noun ending provided gender information (96.1% and 88.3%, respectively). When determiner and noun ending gender information conflicted, participants continued to rely on noun endings over determiners to accurately assign gender (87.9%). Accuracy was low (36.7%) when no gender information was provided. These results demonstrate the learners' ability to rely on the noun-ending cue.

To summarize the results, Oliphant (1998) found that beginner English learners of Italian were able to assign gender based on endings typically associated with one gender, although they were less accurate on words with ambiguous endings. They were also more likely to rely on the final phoneme as a gender cue, rather than the morphological ending, suggesting that they do not, or at least not yet, engage in "backward processing". Finally, the participants relied on gender information provided in noun endings over gender information provided in determiners. Whereas Karmiloff-Smith (1979), based on her findings with L1 French children, suggests that incorporating syntactic (i.e., determiner) gender cues indicates the development of a more advanced strategy for gender assignment, participants in Oliphant's study seemed unaware of the importance of determiner-noun agreement and used gender information provided in noun endings to assign gender.

In addition to an ability to rely on noun-ending cues to assign gender, learners of French demonstrate an awareness of these cues. Three studies by Hardison (1992) were conducted to investigate L2 gender assignment accuracy and to determine strategies used by L2 learners. In the first study, 81 beginning and 38 intermediate American students of French[12] listened to 46 French nouns read aloud as well as saw the printed words on paper, and were instructed to circle the appropriate article (*le/la*) for each word. The French words had phonological endings that were typical of either masculine or feminine, although words that constitute exceptions (i.e., a word that has a different gender than the one predicted by the phonological ending) were also included. Hardison predicted that the students would incorrectly assign gender to the exception words as a result of relying on the word's phonological ending. The beginner and intermediate students achieved 69% and 75% accuracy, respectively; however, when the exception words were removed from the analysis, the accuracy increased to 75% and 81%, respectively, confirming Hardison's hypothesis that the students were relying on the phonological ending to assign gender.

The second study involved 41 intermediate students enrolled in a third year French conversation class. Following the same procedure as in the first study, the students read

[12] Beginner students were enrolled in either a second or third semester French class; intermediate students were enrolled in a third year phonetics and pronunciation class.

and heard unfamiliar French nouns and circled the appropriate article (*le/la*) for each word. The nouns had phonological endings typical of either masculine or feminine and were categorized as likely to be familiar, less familiar, or unfamiliar to the students. The overall accuracy was 84%; students achieved highest accuracy on the familiar nouns (87%) and lowest accuracy on the unfamiliar nouns (80%). Furthermore, some phonological endings appeared to be strongly associated with one gender over the other, as shown by high accuracy on words with these endings, whereas other endings did not appear to serve as gender cues, based on low accuracy on words with these endings. Hardison (1992) proposed that a potential explanation for the varying associations between phonological ending and gender among these students is the orthographic influence. For example, a noun ending in the phoneme [n] often ends in the letter "e", as in *piscine* (pool), and the final letter "e" often indicates feminine gender. To remove the potential orthographic influence, Hardison conducted a third study in which intermediate French students (in their third year of study) heard 34 infrequent/rare nouns and were asked to determine the gender by circling the appropriate gender (*le/la*) for each word. However, the words were not printed on the answer sheet as in the previous two studies. Therefore, no orthographic cues were available. Overall accuracy was 76%, indicating that the students were able to rely on the phonological endings even in the absence of orthographic cues.

In all three studies, learner strategies for assigning gender to nouns, as described by the participants in writing after assigning gender to each noun, included thinking about what "sounds best", repeating the noun in context (i.e., *du fromage*, [some cheese]), and focusing on the noun ending for sound and spelling cues. Paying attention to the sound of noun endings and visualizing spelling were especially important in the third study. Hardison (1992) concludes that L2 learners use noun ending gender cues to formulate regularities in order to assign gender to French nouns, and that "… the learners are processing cues indicative of gender and utilizing strategies similar to those used by native speakers" (p. 304).

4.3 NNS Gender Agreement Data

The studies described above review data on NNSs' ability to assign gender to French nouns, but do not take into account naturalistic data in which the learners have to carry out gender agreement during production. Holmes and Dejean de la Bâtie (1999) compared gender assignment and gender agreement ability among NNSs of French. Gender assignment was tested with a button-press task and gender agreement with an essay-writing task. Fifty L1 English students who were enrolled in a French university class, and who had been studying French for seven years, first completed the same gender assignment task described in Section 2.2.2. The NNS participants achieved 80% accuracy on the regular nouns, 48% accuracy on the exception nouns, and 75% accuracy on the nonce nouns. The slightly below chance accuracy on the exceptional words and the 75% accuracy on the nonce nouns together indicate that the participants were using the noun endings to determine the words' gender. That is, similar to previous results, the NNSs were able to rely on word ending rules to assign gender to both real words and non-words, but not to the same degree of accuracy as NSs.

Participants also completed two written essays, which were transcribed and coded for determiner and adjective gender agreement accuracy. Overall, the participants achieved higher accuracy on gender agreement (86.8% on determiner agreement and 75.4% on noun-adjective gender agreement) in the written task than on the gender assignment task. The authors speculate that the higher accuracy was due to the participants using words for which they knew the gender. At the same time, the authors point out that inaccurate gender agreement occurred even with common words, such as *jour* (day) and *couleur* (color), leading them to conclude that "many of the foreign language learners lacked accurate gender knowledge even of words they used frequently" (p. 500). However, of the examples of these erroneously gender-marked words provided by the authors, several turn out to be exceptions to gender-typical ending regularities. For example, *monde* (world, masc.) and *pays* (country, masc.) were incorrectly marked as feminine, and *couleur* (color, fem.), *fin* (end, fem.), and *mer* (sea, fem.) were incorrectly marked as masculine. This sample suggests that the learners could, in fact, have been relying on word ending rules during the written task.

While Holmes and Dejean de la Bâtie's (1999) data suggest that NNSs rely on noun ending regularities to carry out gender agreement during a written task, there may be additional factors that affect gender agreement accuracy. One likely factor is the location of the adjective in relation to the noun it is modifying. Bartning (2000) investigated the role of the adjective's location in accurate gender agreement. Specifically, she considered gender agreement in an extended framework of Pienemann's (1998a, 1998b) Processability Hierarchy. Pienemann's Processability Theory (PT) proposes a hierarchy of grammatical structures, based on difficulty, which determines the order of acquisition. That is, a learner's psycholinguistic constraints, or the processing procedures available to the learner, direct the order in which he/she is able to acquire new grammatical structures.

This developmental order cannot be changed; acquisition of a more difficult structure can only begin after the structure below it has emerged in the learner's interlanguage.

Based on Pienemann's (1998a, 1998b) PT, Bartning (2000) outlines a five-level processing hierarchy of gender agreement in French: word (lemma) access (level 1), lexical morphology (level 2), phrasal morphology (level 3), interphrasal morphology (level 4), and main and subordinate clauses (level 5). Only after a learner has added a lemma to his/her lexicon (level 1) can processing of the lemma's grammatical features, such as number and gender (level 2) begin, and so on. Applying this to gender agreement, Bartning suggests that exchange of grammatical information, that is, gender agreement, will first occur at level 2, followed by level 3, etc. Table 3 provides examples of each level.

Table 3

Processing Hierarchy of French Gender Agreement

Level	Example
Level 1: Word (Lemma)	Vert, maison (green, house)
Level 2: Lexical Morphology	Vert-e, maison-s (green-fem., house-pl.)
Level 3: Phrasal Morphology	La/une maison verte (A/the green-fem. house)
Level 4: Interphrasal Morphology	La maison est verte (The house is green-fem.)
Level 5: Main and subordinate clauses	La maison qui est verte (The house that is green-fem.)

Within this framework, Bartning investigated gender acquisition and mastery of the gender agreement system in advanced and pre-advanced L2 learners of French.[13] Using interview data from six advanced Swedish learners of French and interview and oral narration data from nine pre-advanced Swedish learners of French, Bartning evaluated adjective gender agreement in the attributive postposition (level 3) and the predicative position (level 4). In contrast to Pienemann's PT, Bartning's results show that, for advanced learners, adjective agreement in the level 3 attributive position was significantly less accurate (79%) than adjective agreement in the level 4 predicative position (84%), suggesting that gender agreement at level 3 is mastered later than at level 4. In addition, applying feminine gender agreement was more difficult than applying masculine agreement in all positions. According to Bartning, this difficulty is due to the feminine form of most adjectives being "longer, more complex, often irregular, and hence more difficult to produce and automize" (p. 236), an explanation which is consistent with the notion of the feminine form of an adjective being the marked form (i.e., requiring an additional phoneme at the end), and the masculine form being the unmarked, or default, form. However, although the results of the advanced learners are not in accordance with Pienemann's PT, Bartning points out that Pienemann's theory is based on emerging grammatical phenomena, whereas the data from the advanced learners represent later stages of development in which all four levels of gender agreement may have already been mastered. In other words, the advanced learners' lower accuracy at level 3 does not represent the development of these stages, and, therefore, does not falsify the PT. Instead, at the advanced level, it is selecting and applying the correct feminine

[13] The data from the advanced learners are from the same interviews reported on at the beginning of this section (Bartning 2000, Bartning & Schlyter, 2004); the "pre-advanced" are secondary school students with 3 years of French (p. 229).

form, not the exchange of grammatical information throughout a sentence, that poses a problem.

For the pre-advanced learners, who are more likely to be in the emerging stages of development, adjective agreement in the level 4 predicative position (77%) was less accurate than agreement in the level 3 attributive position (82.5%). These results are consistent with Pienemann's (1998a, 1998b) PT. As for advanced learners, feminine agreement was more difficult than masculine agreement for pre-advanced learners in all positions, and looking just at feminine adjectives, agreement accuracy in the level 3 attributive position was lower (36%) than in the level 4 predicative position (50%). Bartning (2000) concludes that in gender agreement, choosing the correct gender form is more difficult than exchanging gender information throughout the sentence. That is, learners have more difficulty with morphological form, specifically the feminine form, than exchange of grammatical information across clause boundaries.

Bartning's (2000) findings are replicated by Dewaele and Véronique (2001), who also found gender agreement results that are contradictory to Pienemann's (1998a, 1998b) PT, suggesting that at the advanced level, once all levels of the PT are mastered, it is the processing of the feminine form that accounts for gender agreement errors, not the exchange of grammatical information. Similar to Bartning, Dewaele and Véronique conducted a study that investigated gender agreement under Levelt's (1989) model of speech production and Pienemann's Processing Hierarchy Hypothesis. They addressed several questions, of which the following two are most relevant to this discussion: do constituent borders affect the transfer of grammatical information, and is there a relationship between gender accuracy and the modifier's grammatical class (determiner vs. adjective)?

Trilingual Dutch-French-English university students who had been enrolled in French courses for 4-6 years and were currently enrolled in a university-level intensive French class[14] were informally interviewed and given an oral proficiency test in a formal situation; conversations were recorded and coded at the word level. Participants' adjective agreement accuracy in the level 3 attributive position (89.5%) was not significantly different from adjective agreement accuracy in the level 4 predicative position (92%). The breakdown between errors on masculine and feminine nouns shows that feminine adjective agreement was less accurate than masculine adjective agreement. Of the adjective agreement errors produced, 63% were the masculine used to modify a feminine noun. These results are consistent with those of Bartning (2000) and suggest that for advanced learners, transferring diacritic information across constituent boundaries does not result in lower gender agreement accuracy. That is, even though pre-advanced learners' gender agreement errors follow Pienemann's (1998a, 1998b) PT, gender agreement problems persist through the advanced level at all levels of the processing hierarchy. These errors no longer represent a developmental process, but rather indicate processing constraints of a different nature.

Further evidence for the role of processing constraints comes from a second study by Dewaele and Véronique (2000). They investigated the relationship between mastery of different morphological systems, specifically comparing gender agreement to other types of agreement errors (number agreement and subject-verb agreement), fluency,

[14] These 27 participants are the same as those described in Dewaele & Véronique (2001) at the beginning of this chapter.

complexity, lexical richness, and style choice in advanced French L2 interlanguage. Dewaele and Véronique recorded advanced university students' conversations under two conditions, neutral and stressful, and then transcribed and coded at the word level for analysis. Although accuracy rates are not provided in the original article, the authors present the following findings: determiner agreement accuracy was higher than adjective agreement accuracy; number and subject-verb agreement accuracy was higher than gender agreement accuracy, with no correlation between them; gender agreement accuracy depended on the situation and the fluency of the speaker, with higher accuracy in the informal, neutral condition than the formal, stressful condition; and lastly, speakers with higher fluency (i.e., faster rate of speech) made fewer gender agreement errors.

Based on these results, the authors draw several conclusions. First, because determiner agreement has a higher accuracy rate than adjective agreement, the learners are correctly assigning gender at the lemma level. That is, adjective errors do not signify gender assignment errors because gender was correctly assigned for the determiner. One explanation for this finding is that determiners are frequent, non-complex words whereas adjectives are lower frequency and, as Bartning (2000) pointed out, more morphologically complex. Second, the fact that number and person agreement errors are lower than and do not correlate with gender agreement errors suggests that such errors are not a result of exchanging grammatical information throughout the sentence. Because number and person agreement also requires the exchange of grammatical information throughout a sentence, and this exchange does not pose a problem, it is logical to presume that the exchange of gender information throughout a sentence is not problematic. Rather, gender agreement errors are likely to be the result of processing difficulties. This conclusion is supported by the finding that gender agreement accuracy is dependent on both interview context and speaker fluency. Dewaele and Véronique (2000) explain that learners have quicker access to words they are more familiar with (which are more likely to be used in an informal situation) than words that are not part of their usual discourse domain (which may be employed in a formal situation), and relying on and processing phonological cues becomes more difficult in a stressful situation. Also, more fluent speakers are faster at processing to begin with, and are consequently better able to maintain gender accuracy under formal situations. Therefore, the notion that gender agreement errors are a result of processing problems is supported by the higher error rate in formal situations and among less fluent speakers.

To summarize this chapter, the studies reviewed demonstrate that, despite an ability to reliably assign gender based on noun ending regularities, L2 learners of French are much less consistent in realizing gender agreement throughout a sentence. This pattern suggests that, although NNSs assign gender in a similar manner as NSs, the way in which gender is processed is different. De Bot's (1992) model of bilingual language production, which specifically addresses aspects of L1 processing that are shared with or separate from the L2, offers a potential explanation as to why NNSs are unable to achieve native-like gender agreement, and is presented in Chapter 5.

Chapter 5: Differences in L1 and L2 Gender Representation and Processing

5.1 Model of Bilingual Language Production

Expanding on Levelt's (1989) model, De Bot (1992) describes a model of speech production for a bilingual speaker that accounts for (1) the separation or mixing of two or more languages, (2) cross-linguistic influences, (3) sufficient capacity to maintain production speed, (4) unbalanced proficiency levels, and (5) unlimited number of languages with varying typological distance. De Bot proposes that Levelt's Conceptualizer has both language-nonspecific and language-specific phases, such that the preverbal message contains information about the language(s) to be used during production, and, particularly relevant to this dissertation research, the Formulator contains separate processing systems for each language, but a shared lexical store. That is, a lemma is selected from a common store, but the grammatical information for that lemma is activated via language-specific processing procedures. De Bot claims that although the L1 and L2 share a common lemma store, the two languages rely on separate grammatical encoding processes. In other words, the processes used for carrying out grammatical gender agreement in the L1 are different from the processes used in the L2.

The following sections present studies that examine differences in NS and NNS gender representation and processing, specifically within the framework of the gender storage and nodal model presented above, and consistent with De Bot's (1992) model of bilingual language production. Sections 5.2 and 5.3 address NS-NNS differences for NNSs whose L1 is English, which contains a minimal gender system compared to the L2 (French and German), and Section 5.4 addresses the role of L1-L2 gender-system similarity (German and Romance L1 learners of Dutch) in gender processing.

5.2 NS and NNS Gender Congruency Effects

Guillelmon and Grosjean (2001) looked at how NSs and early and late English-French bilinguals react to gender marking in a French word-naming task. Participants were aurally presented with determiner-adjective-noun phrases in which the determiner was either congruent/incongruent (*le/la*) or neutral (*leur*) with the noun, and the adjective was phonologically neutral in oral speech (e.g., *joli/jolie*). The participants were asked to repeat the noun as quickly as possible. The NSs and early bilinguals showed facilitation and inhibition effects based on the congruency condition; however, the late bilinguals did not. This difference is not likely to be a result of different proficiency levels because all the late bilinguals, who began acquiring French around age 15, moved to a French-speaking country in early adulthood and had been active bilinguals for up to 24 years; they also self-reported proficiency levels similar to those of the early bilinguals. Nor is the inhibition effect in the NSs and early bilinguals due to effects of ungrammaticality, because facilitation effects were found when compared to the neutral condition. The authors speculate that perhaps the late bilinguals "have not established any gender connections among the words sharing the same gender or that they have not given a gender feature to the nouns" (p. 509). However, the participants performed well (32.5/36) on a follow-up gender assignment task, which suggests they do have these connections, but do not activate them during (auditory) processing. An alternative explanation proposed by the authors is that there is a syntactic module that checks the gender agreement, but the late bilinguals have either not developed this "mechanism" or they are unable to use it during perception. Either way, the results suggest that for late bilinguals, word recognition does not automatically activate a grammatical gender node as it does for NSs.

5.3 NS and NNS Reliance on Phonological Cues

Additional evidence that NSs, but not NNSs, possess a nodal relationship between determiners and nouns comes from a study that examined the influence of nouns' lexical and sublexical information in production and recognition tasks (Holmes & Segui, 2006). Lexical information refers to determiners associated with a noun's gender and which provide gender information when the noun is consonant-initial (*le/la*), but not when the noun is vowel-initial (*l'*); sublexical information refers to word endings that are generally associated with one gender over the other. The nouns used in the study were categorized into four conditions: consonant-initial, vowel-initial, ending-typical and ending-neutral. NSs and advanced learners of French[15] first completed an implicit production task in which they were familiarized with English-French word pairs, and were then asked to evoke mentally the French translation of an English word and make a semantic (concrete/abstract) or gender (masculine/feminine) classification decision. In order to separate the translation RT from the classification RT, participants did not know which classification (semantic or gender) they would be asked to make until after they had evoked the word, as indicated by a button push that triggered a classification prompt. They then completed a recognition task in which they were shown printed French words and asked to make a gender classification (masculine or feminine).

In the implicit production task, NSs showed no difference in gender classification accuracy among any of the conditions. They were slower, however, to classify the vowel-

[15] The majority (17) of the learners of French were "advanced undergraduate students of French" and 8 were 1st or 2nd year students of French. However, no proficiency measure is included.

initial words than consonant-initial words. There was no RT difference for gender-neutral vs. gender-typical endings. The NNSs were more accurate at classifying consonant-initial words than vowel-initial words, and there was a trend toward faster classification of consonant-initial words than vowel-initial words. NNSs were also more accurate and faster at classifying words with gender-typical endings than words with gender-neutral endings.

In the recognition task, NSs had more difficulty classifying words (RT and accuracy) when the word was both vowel-initial and ending-neutral. The NNSs' performance was inferior, though similarly influenced. However, both the implicit production and recognition tasks reported in this study may rely more on meta-linguistic knowledge than gender-nodal activation, making it difficult to determine the role of lexical and sublexical information in the gender storage and nodal system. At most, the finding that NSs did not show effects of noun ending in the implicit production task, but were negatively affected by vowel-initial words, provides support for a nodal relationship between the determiner and word. That is, activation of a noun's gender will be strengthened by the simultaneous activation of the noun's associated determiner. When a word activates a determiner that does not indicate gender, such as *l'* in the case of vowel-initial words, activation of the word's gender does not receive the extra boost that occurs when a word activates a determiner that does carry gender information, such as *le/la*. If this is the case, then the results of this study show that NNSs do not have access to this relationship during production, or at least not to the same degree as NSs. Or, it may be that they are paying too much attention to noun endings, and thereby not able to take advantage of nodal activation.

Whereas NSs possess a nodal relationship between the determiner and noun, evidence that NNSs instead rely on gender-marked noun endings during processing has been shown by a study with German NSs and learners of German. Bordag et al. (2006) examined the role of a noun's gender-marked phonological ending on L1 and L2 gender processing. German NSs and English intermediate/low advanced learners of German completed two tasks in German: a picture naming task of nouns with typical, ambiguous or atypical gender-marked endings, and a grammaticality judgment of gender- congruent and incongruent noun phrases.

In the picture naming task, participants were shown 48 line drawings of concrete, mono-morphemic nouns. One third of the nouns had phonological endings typical of the noun's gender, one third had phonological endings that were ambiguous and not associated with a particular gender, and one third had phonological endings that were atypical of the noun's gender. The pictures were presented in two separate conditions to all participants. In the short condition, the picture appeared on the right side of a computer screen and participants were instructed to provide the name of the noun as quickly and as accurately as possible. In the long condition, two versions of the same picture appeared on the screen. The one to be named appeared on the right side of the screen, and either a larger or smaller version of the same picture appeared on the left side of the screen. The participant was instructed to name the picture in German along with the appropriate gender-marked adjective, "big" or "small".

In the grammaticality judgment task, participants were presented with NPs (demonstrative pronoun + noun) and asked to determine by the press of a button whether the phrase was accurate or inaccurate. The NPs consisted of pronouns whose gender was either congruent or incongruent with that of the noun, and within each condition, the nouns had gender-marked endings that were typical, atypical, or neutral.

The NSs' accuracy and RTs in the picture naming task were not influenced by the noun's ending. That is, in both the short (bare noun) and long (adjective + noun) conditions, they were equally accurate and fast at naming the nouns, regardless of whether the noun had a typical, ambiguous, or atypical gender-marked ending. However, the NNSs showed differences among the three types of gender-marked endings in the long condition. They were slowest and made the most errors on nouns with atypical endings, followed by the ambiguous condition. They were fastest and made the fewest errors on the gender typical condition.

The grammaticality judgment task showed the same pattern of results as the picture naming task. The NSs were equally accurate and fast at judging the NPs, regardless of whether the noun ending was typical, atypical, or ambiguous. The NNSs, on the other hand, were faster and more accurate at judging NPs in the typical condition and slowest and least accurate at judging NPs in the atypical condition.

The authors review two possible explanations for their results. First, NSs and NNSs process gender differently; NSs acquire and store grammatical gender as an inherent part of the noun, a feat that is not possible for NNSs, who compensate by relying on phonological or other (e.g., L1 gender) information. Alternatively, NSs and NNSs process grammatical gender in a similar way, but are at different stages of the learning curve. That is, adult learners rely on phonology in the same way NS children do, but with more experience, NSs have mastered gender and a reliance on the phonology is no longer necessary.

To summarize the findings thus far, gender congruency effects in NSs and gender-marked word ending effects in NNSs provide evidence that NSs rely on a gender-nodal system, whereas NNSs do not seem to have access to this system, or at least not to the same degree as NSs. While these findings support De Bot's (1992) claim that learners do not use L1 processing procedures for the L2, these studies only consider learners whose L1, English, does not have a grammatical gender system. That is, these learners do not have an L1 processing procedure specific to grammatical gender agreement that could be used for processing the L2. Research by Sabourin and Stowe (2008) addresses the role of the L1 gender system in L2 grammatical gender processing. If learners of an L1 with a similar gender system to that of the L2 are unable to achieve native-like proficiency in L2 gender agreement, then it suggests they are not able to use their L1 gender agreement processing procedures for the L2, thus, supporting De Bot's claim. If, on the other hand, these learners are native-like in L2 agreement, it suggests that they may be taking advantage of their L1 processing procedures.

5.4 Role of the L1 in L2 Gender Processing

Sabourin and Stowe (2008) tested 23 NSs of Dutch, 14 German L1 and 8 Romance L1 advanced learners of Dutch on a grammaticality judgment task. Both German and the Romance languages (French, Italian, Spanish) included in the study have grammatical gender systems that are similar to that of Dutch in that nouns contain grammatical gender, which is marked on determiners. However, the Dutch system has a closer correspondence to German than to the Romance languages at the lexical level. For example, German neuter translates to Dutch neuter, and German masculine and feminine translate to Dutch common gender. French, Italian, and Spanish, on the other hand, have a one-to-many correspondence because masculine and feminine may translate to either Dutch neuter or Dutch common genders.

The grammaticality judgment task consisted of 40 target sentences with determiner-noun gender agreement errors and their grammatical counterparts. The sentences were presented one word at a time in the middle of a computer screen, each word appearing for 250ms. The gender agreement error became evident at presentation of the noun, which occurred in the middle of the sentence or in the sentence final position. Participants were instructed to judge the sentence as either grammatical or ungrammatical by button push at the end of the sentence. Both accuracy and ERPs were recorded. A significant relationship between L1 and accuracy was found, with the NSs performing significantly better (94.3% accuracy) than both NNS groups, and the German L1 participants performing significantly better (82.1% accuracy) than the Romance L1 participants (59.4% accuracy). The NNS non-native-like accuracy is not due to the participants not knowing the gender of the target nouns, as their follow-up gender assignment task accuracy was 93% (German L1) and 78% (Romance L1). Interestingly, despite their non-native accuracy, the German L1 participants showed a similar P600 effect to that of the Dutch NSs, indicating a sensitivity to the gender violation, although the magnitude was not as great and the peak occurred later than for the NSs. The Romance L1 participants did not show a P600 effect; however, they did show an early frontal negativity similar to that of the Dutch NSs. It seems that the near perfect correspondence between the German and Dutch gender systems was advantageous for the German L1 participants. Regarding the Romance L1 participants' frontal negativity, the authors speculate it is related to "a memory resource that is used when attempting to maintain information in the hopes that a resolution for the ungrammaticality will be encountered later in the input…" (p. 424). They continue on to suggest that WM may be the resource used by these learners who "have some awareness of the ungrammaticality but are not able to use native processing routines to deal with the ungrammaticality" (p. 424). Similar frontal negativity was found and attributed to the use of a memory resource in Sabourin and Stowe (2004); in this study, the effect was related to the ungrammatical word occurring mid-sentence, as opposed to in the sentence final position, for which no frontal negativity was found. The participants had to consciously maintain the fact that an error had occurred until the end of the sentence, when their judgment could be made, thus, creating a memory load that resulted in a frontal negativity ERP effect. Sabourin and Stowe's (2004) findings add support to their supposition that a memory component plays a role in maintaining gender information throughout the sentence, at least for the participants for whom the L1 gender system is not congruent to the L2 gender system.

To summarize the results from Sabourin and Stowe (2008), neither the German nor Romance L1 participants achieved native-like accuracy on the grammaticality judgment task. The German L1 participants showed similar P600 effects to those of the Dutch NSs, but the Romance L1 participants did not, although they did show a similar frontal negativity that the authors interpret as use of a memory resource. It is important to note, however, that despite German L1 near-native-like P600 effects, P600 effects alone do not imply native-like processing. Tokowicz and MacWhinney (2005) found P600 effects in English L1 beginner learners of Spanish on grammaticality judgment task sentences containing gender agreement errors, despite grammatical judgment accuracy at chance. The sentences contained determiner-noun gender agreement errors, and were presented one word at a time in the middle of a computer screen. Participants made grammaticality judgments at the end of the sentence by pushing one of two buttons, and ERPs were recorded. Clear P600 effects were found on the ungrammatical sentences, although the participants' accuracy was 36%, which, when taking into account the yes-bias, was at-chance performance. Therefore, in the case of Tokowicz and MacWhinney, the native-like P600 effect indicates sensitivity to the gender violations, but does not indicate native-like gender agreement ability.[16] Although the authors conclude that this ERP sensitivity is due to implicit syntactic processing, the low accuracy is far from native-like, and one cannot conclude that the NNSs demonstrated native-like processing. In other words, that P600 effects exist even in beginner L2 learners who are unable to accurately determine a sentence's grammaticality implies that P600s do not equate to native-like processing. Therefore, although the German L1 participants in Sabourin and Stowe's (2008) study are very similar to NSs in their ERP effects, this is not sufficient evidence that they are native-like.

5.5 Summary of L2 Gender Processing

Overall, Sabourin and Stowe's (2008) findings support De Bot's (1992) model in that NNSs with similar L1-L2 gender systems did not achieve native-like accuracy, nor did they show native-like processing (although it was very similar for the German L1 participants) on an on-line L2 gender agreement task. That is, the NNSs of German were not using their L1 gender agreement processing procedures for L2 gender processing. However, the findings did suggest that the closer the relationship between the L1 and L2 gender systems, the greater the advantage in terms of processing gender in a native-like

[16] It is important to note that this study did not include NS controls, so the "native-like" P600s is referring to P600s found in NSs in other studies, such as Sabourin & Stowe (2008).

fashion. The role of L1-L2 transfer will be addressed in more detail in the next chapter. In the case in which the L1 and L2 systems are not similar, or the L2 system is unique, NNSs may rely on surface cues, such as phonological and morphological noun endings that provide cues as to the noun's gender. Furthermore, the Romance L1 participants' ERP effects suggest that a memory component may play a role in non-native gender agreement processing. The potential role of a memory component is compatible with the L2 naturalistic data described above, especially Dewaele and Véronique's (2000) conclusion that gender agreement errors are linked to processing difficulties. Specifically, that determiner-noun gender agreement, in which the determiner most often directly precedes the noun, is consistently higher than adjective-noun agreement, in which the adjective may occur several words down from the noun it modifies, suggests that a memory component is involved. This postulation is also raised by Holmes and Dejean de la Bâtie (1999), who suggest "an additional memory component is introduced by the fact that most adjectives appear after the noun, sometimes appearing several words downstream in predicate adjective constructions" (p. 500).

It is proposed in this dissertation research, therefore, that WM plays an important role in holding a noun's gender information in memory while continuing to process the remainder of the sentence in order to carry out appropriate gender agreement. Moreover, the role of WM is hypothesized to be especially relevant for NNSs who do not have a similar gender system in the L1. That is, with no possibility of transferring L1 processing procedures, these learners will rely on non-linguistic processes, such as WM.

5.6 Grand Summary

Chapters 2 through 5 provided an overview of studies on L1 and L2 grammatical gender representation and processing; this section provides a general summary of the findings.

As children, NSs of French use phonological, morphological, and semantic rules to acquire the gender classification system and apply it to assigning gender to new words. However, once a word is acquired, NSs do not use gender cues during gender agreement. Instead, grammatical gender is stored as syntactic information at the lemma level and is automatically activated, prior to and independent from the phonological form, when a lemma is selected. NNSs, however, do not seem to represent gender information similarly to NSs, with the main differences being that NNSs do not benefit from automatic gender activation when a lemma is selected, but do benefit from phonological and morphological gender cues, as demonstrated in gender congruency tasks. NNSs also do not appear to process gender similarly to NSs, as indicated by nonnative-like realization of gender agreement, both in gender agreement accuracy scores and physiological responses to gender violations. However, there is some evidence that L1-L2 gender-system similarity may provide an advantage during gender processing, and a memory component may play a role as well.

No study to date has addressed both L2 gender representation and gender processing while taking into consideration the role of the L1 and other potential factors, such as gender cues and WM, that may contribute to NNS gender agreement ability. The current study attempts, therefore, to investigate the differences between L1 and L2 gender representation and processing, as well as the role of the L1, gender cues, and WM in L2 gender processing.

Chapter 6: Current Study

This study investigates how NNSs represent grammatical gender and realize[17] gender agreement in spoken French, and how their L1s, gender cues, and non-linguistic processing constraints affect their development in both aspects. A gender priming task and a grammaticality judgment task are used to investigate gender representation and gender agreement, respectively. Because the naturalistic data (Dewaele & Véronique, 2000, 2001; Holmes & Dejean de la Bâtie, 1999) suggest that determiner-noun agreement accuracy may reflect incorrect gender assignment rather than gender agreement processing difficulty, and because one of the goals is to examine the role of WM, only noun-adjective agreement, and not determiner-noun agreement, were considered in the grammaticality judgment task designed to examine NNS gender agreement. Examining noun-adjective agreement allowed for manipulating the location of the adjective in relation to the noun it modifies. In order to examine the role of the L1 in acquiring an L2 gender system, three L1 groups were considered: Spanish, whose gender system is ostensibly congruent to that of French; Dutch, whose gender system is incongruent to that of French; and English, whose gender system is minimal relative to French. An overview of Spanish, Dutch, and English gender agreement systems will be provided below in order to illustrate how they differ from French.

According to the gender storage and nodal model and De Bot's (1992) model, an L2 learner has to develop a grammatical information storage system from which L2 processing procedures can pull information in order to produce grammatical utterances.

---

[17] In this dissertation, "realize" refers to the process involved in both recognizing and carrying out gender agreement; however, the experimental tasks only require participants to recognize gender agreement in written French.

Although a NNS may be able to develop a store of gender information to carry out accurate gender agreement, a gender-nodal system is necessary to achieve native-like gender processing.

Specifically for a nodal system to carry out noun-adjective agreement, the learner has to develop a store of grammatical information for each lemma. For French nouns, this would include grammatical gender information that would allow for a lemma-gender node link. The L2 processing procedures would then use this information to coordinate accurate noun-adjective agreement throughout the sentence. Therefore, for a Spanish speaker, the L1 and L2 grammatical information for nouns will both include masculine and feminine gender. For a Dutch speaker, the information will be slightly different in that the L1 includes common and neuter genders, but the L2 will include masculine and feminine genders. For the English learner, only the L2 will require gender information. The Spanish speaker may also use an L2 gender processing procedure that is similar to that of the L1 because realization of noun-adjective gender agreement in Spanish and French is nearly identical. The Dutch speaker, however, will have to develop an L2 gender processing procedure that is different from the L1 procedure because the Dutch gender agreement system is not a close parallel to that of French. Finally, the English speaker will have to develop an L2 gender processing procedure from scratch, as there is no equivalent in the L1.

Although contrastive analysis research has shown that similarities between the L1 and L2 do not necessarily result in ease of learning, and differences between the L1 and L2 do not necessarily create difficulty for a learner (Larsen-Freeman & Long, 1991), the prior research on L1-L2 gender systems presented in the previous chapter shows that similarity does facilitate gender agreement ability and absence of a gender system creates difficulty (Franceschina, 2005; Sabourin & Stowe, 2008, Sabourin, Stowe, & de Haan, 2006). Therefore, it is hypothesized that the Spanish L1 learners of French will have the least difficulty in developing a grammatical information storage system and L2 processing procedures for French gender, and the English learners will have the most difficulty.

Before addressing the specific research questions and predictions of the current study, Pienemann's (1998a, 1998b) Processability Theory and its implications for L1-L2 transfer will be addressed, followed by an overview of the gender systems of the four languages considered in this study.

6.1 Pienemann's Processability Theory and L1-L2 Transfer

Mentioned briefly in Chapter 4, an overview of Pienemann's (1998a, 1998b) PT is presented here and its implications for the current study discussed.

Within the framework of Levelt's (1989) and De Bot's (1992) models, Pienemann maintains that acquiring an L2 involves acquiring the processing procedures specific to the L2, and that the sequence of acquiring these procedures is hierarchical in that "each procedure is a necessary prerequisite for the following procedure" (1998a p. 6). That is, a learner can only process the L2 to the point in the hierarchy that he/she has acquired the processing procedures. Furthermore, Pienemann specifies that these L2 procedures cannot be "bulk transferred" from the L1 (1998a, p. 81), that is, only parts of procedures can be transferred, because even small differences between the L1 and L2 systems would create a processing problem. For example, diacritic features of a lemma (i.e., tense, gender, case) and the destination of this information (i.e., determiner and verb agreement) are different for each language, despite possible overlap; therefore, the L1 processing procedures are not designed to process L2 information. Consistent with De Bot, Pienemann claims that learners have to develop L2-specific procedures, and furthermore, because diacritic features of lemmas are language specific, diacritic information has to be developed separately for the L2. However, although bulk transfer does not occur, transfer of some L1 procedures can occur when the learner is developmentally ready to acquire that procedure. In other words, procedures similar in the L1 and L2 can be acquired through transfer, but only after the processing prerequisites (within the hierarchy) have been developed (1998a, p. 82).

The implications for the current study based on this theory of transfer are that the Spanish and Dutch learners of French will be able to transfer aspects of the L1 gender processing procedures when developmentally ready; the English learners of French, however, will have to develop this procedure. Consequently, if English learners of French must create a new procedure to process French gender, rather than rely on L1-L2 transfer, the question is whether they are able to create a procedure similar to that of a NS. Based on the findings that English learners of French and English learners of German do not represent gender in the same way as NSs of these languages (Bordag et al., 2006; Guillelmon & Grosjean, 2001; Holmes & Segui, 2006), it is unlikely that learners whose L1 does not have a gender system will be able to develop automatic gender processing procedures similar to a NS. They may instead rely on a non-linguistic resource, namely WM to carry out deliberative gender agreement during processing. Furthermore, the availability of external gender cues, such as determiners, should facilitate gender processing for these learners because gender information may not be included in an L2 lemma's store of grammatical information. However, WM and the availability of external gender cues are not expected to facilitate gender agreement processing for Spanish learners of French because their L2 lemmas' store is likely to include gender information, and they may rely on L1-L2 transfer for gender agreement processing procedures. The role of WM and external gender cues for Dutch learners of French is less clear. Although Dutch has a gender system, the rules of gender agreement are not a direct match to those of French. These implications will be addressed in more detail after the following section on the differences between the gender systems of French, Spanish, Dutch, and English.

6.2 Grammatical Gender Systems

As described in Chapter 2, grammatical gender is the division of nouns into classes based on phonological and/or semantic properties, such as sex and animacy (O'Grady & Guzman, 2001). This section provides a brief overview of the gender assignment and agreement systems of the NNSs' languages addressed in this study: Spanish, Dutch, and English. Because this study focuses on noun-adjective gender agreement, and also considers the role of external gender cues in the form of gender-marked determiners, only these aspects of the languages' gender systems are be addressed. Pronouns, which also mark gender in all four languages, are not be included. The patterns of gender assignment in Spanish and Dutch are similar to those of French, and are presented briefly in the sections below. However, they are not considered crucial to L2 gender representation or processing because the ability to use a noun's gender cues to accurately assign gender does not necessarily result in native-like gender representation or processing. That is, even if a Spanish or Dutch NS were able to transfer L1 gender assignment strategies to assign gender to French nouns, previous research, as reviewed in Chapter 4, has shown that native-like gender assignment neither indicates native-like gender representation, nor results in native-like gender agreement during processing. At most, similar L1-L2 gender assignment patterns will facilitate the initial learning of a noun's gender, but as this study considers only highly proficient learners of French, who are likely to have an extensive lexicon with accurate gender knowledge, this potential advantage is no longer relevant.

6.2.1 Spanish

The Spanish gender system is very similar to the French gender system in that masculine and feminine gender is assigned to nouns based on word ending and semantic properties. Using an inverse dictionary (i.e., alphabetized by word-final letter), Teschner and Russell (1984) analyzed gender patterns of Spanish noun endings. They found that nouns ending in [a] and [d] are overwhelmingly feminine (over 90%), and nouns ending in [n], [z], and [s] are ambiguous in that their predictability of one gender over the other is between 40-60%. Nouns ending in the remaining phonemes ([e], [l], [o], and [r], which account for the majority of nouns, and also [i], [m], [t], [u], [x], [y], [b], [c], [tʃ]) are overwhelmingly masculine (over 89%). Although Teschner and Russell do not discuss the role of morphological endings, they do qualify that the two ambiguous phonological endings, [n] and [z], may be predictive of gender when considered within the context of the preceding 1-3 phonemes. Specifically, words ending in -*ción*, -*gión*, -*nión*, -*sión*, -*tión*, and -*xión*, are feminine (although Teschner and Russell cite 13 words with one of these endings as masculine), as well as words ending in -*ez*, but words ending in -*ón*, -*az*, -*oz*, and -*uz* are masculine. Therefore, whereas the hierarchy found in French that determines morphological rules to be dominant over phonological rules does not exist in Spanish; both phonological and morphological noun endings provide reliable cues as to a noun's gender. Finally, as in French, animate nouns referring to humans respect semantic gender, such that words such as woman and girl are feminine, and man and boy are masculine.

Similar to French, gender is marked on definite and indefinite determiners and on adjectives. However, a minor difference is the distinct masculine and feminine forms for plural determiners, for definite and indefinite. Where the difference is neutralized in French, it is not in Spanish, as shown in examples 13-16.

(13) el libro, un libro [masc. sg.]

the book, a book

(14) los libros, unos libros [masc. pl.]

the books, the books

(15) la casa, una casa [fem. sg.]

the house, a house

(16) Las casas, Unas casas [fem. pl.]

the houses, the houses

Similar to French, Spanish adjectives are also marked for masculine and feminine (examples 17 and 18).

(17) El libro es pequeñ<u>o</u>.

The book is small.

(18) La casa es pequeñ<u>a</u>.

The house is small.

Generally, masculine adjectives are marked by an [o] ending, and feminine adjectives by an [a] ending. Adjectives ending in a consonant are typically masculine, with the feminine form created by adding an [a] ending. As in French, some adjectives do not have distinct orthographic or phonological masculine and feminine forms, such as *difícil* (difficult), which has only one form for both masculine and feminine.

6.2.2 Dutch

Whereas French and Spanish both have a masculine-feminine distinction, Dutch nouns are either common gender or neuter gender. Common gender combines nouns that, historically, were either masculine or feminine nouns, although it is important to note that the masculine-feminine distinction is realized in pronouns.[18] The Dutch gender assignment system is also less transparent than that of French and Spanish. A noun's phonological properties do not provide gender cues; however, in some cases, gender may be determined by morphological and semantic properties (Blom, Polišenská, & Unsworth, 2008). According to Blom et al., derivational morphology may provide gender cues; for example, nominalized nouns with the prefix *ge-* (as in *het geloop,* [the walking]) or the suffix -*isme* (as in *het idealisme* [the idealism]) are neuter gender, and nouns ending in the suffixes -*heid* (as in *de waarheid* [the truth]) and -*ine* (as in *de cabine* [the cabin]) are common gender (p. 260). Furthermore, gender may be predictable for nouns falling into semantic classes, such as names of metals and sports, which are neuter gender, and flowers and seasons, which are common gender. Donaldson (1981), in a Dutch reference grammar, provides a list of semantic categories to help students determine a noun's gender, although it is noted that some of the rules are vague and many exceptions exist. Nevertheless, Donaldson cites the following semantic categories:

---

[18] This masculine-feminine distinction in common nouns makes it difficult for NSs to select the appropriate pronoun during speech, as NSs do not always know the gender of a noun, especially if the noun is inanimate and infrequent.

common nouns include animals, trees, flowers, fruit, stones (that are considered as objects), days, months, seasons, mountains, large rivers, musical instruments, and virtues and vices; neuter nouns include minerals, colors, points of a compass, countries, provinces, cites, and villages (p. 27-32). A final category is diminutives, which are always neuter regardless of the gender of the full noun, for example, *de hond* (the dog), but *het hondje* (Blom et al.).[19] Animate nouns referring to humans, such as man (*de man*) and woman (*de vrouw*), are common gender, although their masculine/feminine distinction is relevant for pronoun selection.

Dutch gender is marked on determiners; neutralization occurs in singular indefinite and plural determiners (examples 19-22).

(19) de tafel, een tafel [common sg.]

the table, a table

(20) het huis, een huis [neuter sg.]

the house, a house

(21) de tafels [common pl.]

the tables

(22) de huizen [neuter pl.]

the houses

Gender is also marked on adjectives, but the difference between the common and neuter adjective form is only realized when the accompanying determiner is indefinite singular. In this case, the neuter noun takes an uninflected adjective (that is, without the [schwa] suffix), as shown in examples 23 and 24 below.

---

[19] Donaldson (1981) refers to diminutives as a semantic category; however, they may be considered a morphological process.

(23) de kleine tafel, een kleine tafel, de kleine tafels [common]

the small table, a small table, the small tables

(24) het kleine huis, een klein huis, de kleine huizen [neuter]

the small house, a small house, the small houses

6.2.3 English

Finally, English has only a pronominal gender system in which only third person singular pronouns (he/she) and third person personal pronouns (his/her[s]) mark semantic gender for humans (and animals whose genders are known). Because full nouns do not have gender, there is no gender marking on determiners or adjectives.

Table 4 below presents the determiner and adjective gender marking systems for French, Spanish, Dutch, and English.

Table 4

Overview of French, Spanish, Dutch, and English Gender Systems

		French		Spanish		Dutch		English
		Masculine	Feminine	Masculine	Feminine	Common	Neuter	
Determiners								
Sing:	Definite	Le	La	El	La	De	Het	The
	Indefinite	Un	Une	Un	Una	Een	Een	A
Plural:	Definite	Les	Les	Los	Las	De	De	The
	Indefinite	Des	Des	Unos	Unas	--	--	--
Adjectives								
Sing:	Definite	Petit	Petite	Pequeño	Pequeña	Kleine	Kleine	Small
	Indefinite	Petit	Petite	Pequeño	Pequeña	Kleine	Klein	Small
Plural:	Definite	Petits	Petites	Pequeños	Pequeñas	Kleine	Kleine	Small
	Indefinite	Petits	Petites	Pequeños	Pequeñas	Kleine	Kleine	Small

6.3 Research Questions

Based on the theoretical framework and differing gender systems outlined above, the following research questions were developed:

1. How does the L1 influence French L2 grammatical gender representation?
2. How does the L1 influence French L2 grammatical gender processing?
3. What is the role of external gender cues in French L2 gender processing?
4. What is the role of WM in French L2 gender processing?

Given that the Spanish gender agreement system is similar to the French system, it is expected that the Spanish learners of French will represent and process French gender similarly to French NSs. Specifically, the Spanish learners will be able to create a grammatical gender information store and transfer L1 gender processing procedures to the L2. The lemma store and L1 transfer will allow for a gender storage and nodal system, which, in turn, will result in native-like gender agreement during processing.

The Dutch learners of French should also show gender representation similar to French NSs because creating an L2 grammatical information store that includes gender should not pose a problem. However, because Dutch gender agreement rules are different from those of French, specifically, the common-neuter distinction in adjectives only occurs in indefinite singular, transferring L1 gender processing procedures will not result in native-like gender agreement during processing. Consequently, Dutch learners of French may benefit from external gender cues and also rely on non-linguistic resources, WM, during gender agreement processing.

Finally, the English learners of French, who will be unable to create a store for grammatical gender information because their L1 has no corresponding gender system, will rely on external gender cues during gender agreement processing. In addition, as they have no L1 gender processing procedures to transfer to the L2, they will rely on WM to carry out gender agreement processing. Figure 3 illustrates gender representation for each of the L1 groups. These representations exist at the lemma level, and are used by the gender processing procedures in order to carry out gender agreement.

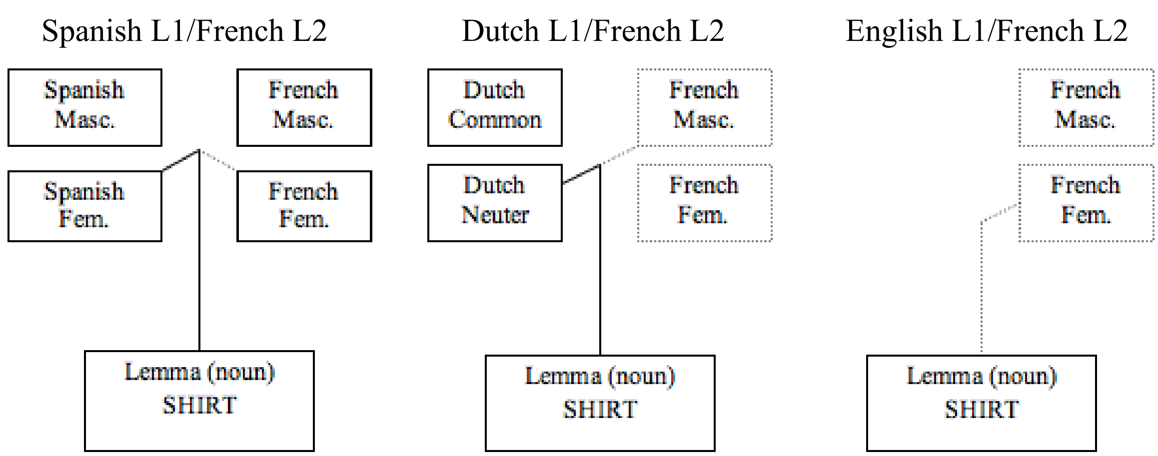


Figure 3. Illustration of L2 French gender representation for Spanish, Dutch, and English NSs

Each L1 learner of French has to create an L2-grammatical-information store for each lemma. For Spanish L1 learners of French, the L2-grammatical information is identical to that of the L1, thus, the masculine and feminine gender nodes are easily created. In addition, a lemma-gender nodal link already exists, and must only be adjusted to connect to the appropriate L2 gender, as indicated by the dotted line to the French feminine gender node. The Dutch L1 learner of French, however, has to create L2 grammatical information that is different from that of the L1. The dotted lines around the masculine and feminine gender nodes represent the new and different L2 gender nodes that must be created. Because the Dutch L1 learner, like the Spanish L1 learner, already has lemma-gender nodal links established, the link must be adjusted to the appropriate L2 gender node. The English L1 learners of French must create gender nodes from scratch, which is indicated by the dotted lines around the gender nodes. In addition, because the English L1 learner does not have lemma-gender nodal links, these links must also be created from scratch, as indicated by the dotted lines from the lemma to the L2 gender node. Based on these models in which the solid black lines indicate what already exists in the L1, and the dotted lines indicate what does not exist in the L1, it is clear that the Spanish L1 learners of French will be able to transfer the most from the L1 to the L2, and the English L1 learners the least.

The following general hypotheses summarize these predictions. More specific hypotheses will be presented after each experimental task description in the next chapter.

1. Spanish and Dutch, but not English, learners of French will represent grammatical gender similarly to French NSs.
2. Spanish, but not Dutch or English, learners of French will realize gender agreement similarly to French NSs.
3. Dutch and English, but not Spanish, learners of French will rely on external gender cues during gender processing.
4. For Dutch and English, but not Spanish, learners of French, WM span will be correlated with accurate gender agreement.

Chapter 7: Methodology

This chapter describes the participants, experimental tasks, and general procedure for this study. For the three main tasks - gender priming, grammaticality judgment, and operation span - the materials, experimental design and procedure, and pilot study results are presented, followed by brief descriptions of the gender assignment post-test, language history questionnaire, and general procedure. Specific details regarding the experimental procedure and logistics are presented in Appendix H.

7.1 Participants

To investigate the hypotheses presented in Chapter 6, it was necessary to examine gender processing patterns of advanced L2 learners of French who (a) have a thorough understanding of the French gender agreement system, and (b) have mastered all levels of Pienemann's (1998a, 1998b) PT. That is, advanced learners are assumed to have mastered complex sentence structures and are able to produce and comprehend such sentences without exceptional processing strain. In addition, advanced learners will have had the opportunity to transfer L1 processing procedures to the extent possible. With these variables controlled, gender agreement errors can be assumed to be solely the result of the L2 learner gender agreement system, and not representative of interference from other aspects of the learner's interlanguage.

To ensure that the participants in this study were indeed advanced learners of French, only students enrolled in a French graduate program either in France or Belgium, or working professionals living in France at the time of data collection, were eligible to participate in the experiment. Furthermore, only adult learners of French were considered. Because child language learners are rarely unsuccessful at achieving native-like proficiency in an L2 (see Hyltenstam & Abrahamsson, 2003, for a review of age effects in SLA), an early childhood learner of French (i.e., age 0-8) would not be expected to have difficulty mastering the French gender system; such a learner, therefore, would not contribute to our understanding of the challenges an L2 learner faces in acquiring the French gender system. Finally, in an attempt to minimize the influence of an L3 gender system, only participants who had not studied a Romance language other than French for more than three years were recruited to participate.

A total of 138 participants participated in the main study: 21 French NSs (FNSs), 37 Spanish NSs (SNSs), 38 Dutch NSs (DNSs), and 42 English NSs (ENSs). An additional 13 participants participated, but were excluded for the following reasons: three (one FNS, one SNS, and one ENS) were excluded due to a technical malfunction during the gender priming task; six (three SNSs and three DNSs) were excluded due to picture naming accuracy scores below 40% on the gender priming task; two SNSs were excluded because it became apparent in the language history questionnaire that they began learning French at the ages of 6 and 7 in a French immersion school; one ENS was excluded because it became apparent in the language history questionnaire that the participant was an English-Spanish early bilingual; and one FNS was excluded because the grammaticality judgment task was not properly completed. Profiles of the 138 participants who were included in the study will be presented in Chapter 8.

7.2 Pre-screening Task

All potential participants completed a web-based grammaticality judgment task prior to participating in the experiment to assess their proficiency level. To create the pre-screening task, 24 French sentences were written, with half of the sentences containing no errors and half containing errors on complex negation (*rien*, *aucun*), verb form (*avoir* vs. *être*), or subject-verb agreement (**ils prend* vs. *ils prennent*). According to Bartning and Schlyter's (2004) work on stages of development in learners of French, errors of these types disappear in the advanced stages of acquisition. All sentences were carefully reviewed by a professor of French. After revisions were made, eight French NSs reviewed the sentences in order to ensure that the sentences were native-like and there were no errors other than the ones intended.

To determine an appropriate cutoff for excluding learners from participating in the main study, the pre-screening sentences were piloted with nine NNSs of French, six of whom were highly proficient (two professional translator/interpreters, a French immersion school teacher, two PhD candidates in French literature, and one undergraduate who has spent extensive time living in France), and three of whom had not yet reached high proficiency (three undergraduate French majors who had all spent at least one semester abroad), termed low proficiency here for convenience. In addition, three of the six high proficiency NNSs and all three low proficiency NNSs had participated in a previous experiment conducted by the researcher; accuracy scores from two tasks in the previous experiment were used to confirm that the high proficiency NNSs were indeed higher proficiency than the low proficiency NNSs. In the previous experiment,[20] these six NNSs completed a translation and picture naming task, based on Kroll and Stewart (1994), in which they were required to translate 30 English words into French and name 30 pictures in French as quickly and as accurately as possible. The

[20] This experiment refers to the researcher's qualifying paper, which served as a pilot study for the current dissertation.

learners also completed a grammaticality judgment task in which they heard 240 French sentences read by a NS and were required to determine whether the sentences were grammatical or ungrammatical.[21] Accuracy scores on the translation and picture naming task and the grammaticality judgment task filler sentences confirmed that the three high proficiency NNSs performed better than the three low proficiency NNSs. The mean accuracy scores of the three high proficiency NNSs were 82% for the translation and picture naming task and 86% for the grammaticality judgment task fillers, and the mean accuracy scores for the low proficiency NNSs were 72% for the translation and picture naming task and 70% for the grammaticality judgment task fillers. In order to use the pre-screening sentences to determine who qualifies to participate in the current study, the sentences needed to discriminate between those in the high proficiency group and those in the low proficiency group.

To this end, an item analysis was carried out for each of the 24 sentences. First, the item difficulty for each sentence was calculated by dividing the sum of correct responses by the number of participants. This was done separately for the two groups (high proficiency and low proficiency). A score of 1.0 indicates that all participants judged the item accurately, and a score of 0.0 indicates that all participants judged the item incorrectly. Using these scores, the item discriminability for each sentence was calculated by subtracting the item difficulty for the high proficiency group from the item difficulty for the low proficiency group. This score determined whether the sentence was distinguishing between the high and low proficiency NNSs; a score of 1.0 indicates that

[21] This task included 80 incorrect target sentences, all of which contained noun-adjective gender agreement errors, and 160 filler sentences, 120 of which contained no errors, and 40 of which contained verb agreement or adverb placement errors. However, in order to use a measure independent from the target of this study (gender agreement), only the filler sentences were used to select participants to pilot the pre-screening task.

all high proficiency participants judged the item correctly and all low proficiency participants judged the item incorrectly; in other words, the item is properly discriminating between the high and low proficiency groups. On the other hand, a score of -.33 indicates that only half of the high proficiency participants judged an item correctly and all of the low proficiency participants judged the item correctly. In this case, the low proficiency participants performed better on the item than the high proficiency participants. There were four sentences on which the low proficiency NNSs performed better than the high proficiency NNSs; these sentences were excluded. Of the remaining 20 sentences, 11 had an item discriminability of 0.0, meaning both the low and high proficiency groups performed equally well. In order to maintain an equal number of correct and incorrect sentences in this task, five of these 11 sentences were selected to be included in the task, and six were excluded. The mean item discriminability for the final 14 sentences was .24 (range .00-1.00). Finally, the mean accuracy score for the high proficiency and low proficiency NNSs was calculated; the high proficiency group scored 90% (range 86%-100%) and the low proficiency group scored 67% (range 64%-71%). These results indicate that the pre-screening sentences are discriminating between highly proficient learners of French and learners of French who are advanced, but not highly proficient. Furthermore, an 85% cutoff was deemed appropriate for selecting only highly advanced learners of French to participate in this study. Therefore, participants who judged more than two sentences incorrectly did not qualify for the main experiment. In order to ensure that participants were judging incorrect sentences based on the intended errors, participants were asked to provide a correction to the sentences they judged as incorrect. This allowed the researcher to assess whether the participant was able to accurately identify *and* correct the error. Sentences that were accurately judged as incorrect, but were not accurately corrected, were counted as errors. For example, if a participant correctly judged the following sentence as incorrect, **Grâce à la gentillesse de l'infirmière, les malades reprend de courage*, but entered the correction as **les malades reprendend*, the participant did not receive credit. Requiring participants to correct the error also prevented participants from losing points if they incorrectly identified a correct sentence as incorrect based on a stylistic or punctuation preference. For example, if a participant judged the following sentence as incorrect, *La petite fille adorait sa poupée, donc quand elle l'a perdue, elle était vraiment triste*, but entered the stylistic correction *sa poupée; donc*, the participant received credit. The final pre-screening sentences, along with their item discriminability scores, are included in Appendix A, and the detailed procedure of the pre-screening task is presented in Appendix H.

A total of 167 potential participants completed the pre-screening task; 139 scored above the 85% cutoff, and, therefore, qualified to participate in the main experiment, and 28 scored below the 85% cutoff. The mean percent correct of those who qualified was 93.6% (range 85.7%-100%); these participants were invited to participate in the main experiment.

Finally, as described above, an attempt was made to minimize the influence of an L3 gender system by only recruiting participants who had not studied a Romance language other than French for more than three years. However, it is difficult to find highly proficient learners of French who have not also studied another Romance language. While most participants met this requirement, 13 participants (one SNS, six DNSs, and six ENSs) had studied another Romance language for more than three years. However, all had begun learning French earlier and/or had been studying French longer than the other Romance language, and furthermore, all were either immersed in French or using French regularly in their graduate program at the time of the study.

7.3 Gender Priming Task

To address the first research question, whether NNSs store grammatical gender information as an inherent property of the noun, a gender priming task was developed. Following Alario et al. (2004), participants were presented with a gender prime, followed by a target picture. Participants were asked to name the picture in French as quickly and as accurately as possible.

7.3.1 Materials

The target pictures were 48 nouns (24 feminine, 24 masculine) and their corresponding line drawings taken from Snodgrass and Vanderwart (1980). The stimuli for this task were selected based on their having been normed for name agreement, image agreement, conceptual familiarity, and visual complexity for English NSs (Snodgrass & Vanderwart), and more recently, for French NSs (Alario & Ferrand, 1999). Only words with 93% -100% name agreement for French NSs (Alario & Ferrand) were included.

Frequency for the target nouns was taken from the www.lexique.org database (New, Pallier, Ferrand, & Matos, 2001) and ranged from 1.44 - 504.15 per million (mean 56.89). The range of frequencies, though not considered as an independent variable, may yield different patterns among the L1 groups and provide some insight as to whether gender information can be stored inherently for NNSs.

The target nouns included both ambiguous and unambiguous phonological endings, based on Surridge (1993, 1995) and Lyster (2006). Because Surridge's and Lyster's classifications are not identical, the following criteria were used to determine an ending's ambiguity:

1. If both authors determined an ending to have above 70% predictability for the same gender, it was considered unambiguous.
2. If a phonological ending was determined have gender predictability less than 70% by one author, but predictability for the (same) gender above 70% by the other author, it was considered unambiguous.
3. If an ending was determined to be less than 70% predictable by both authors, it was considered ambiguous.
4. If an ending was determined to be feminine by one author but masculine by the other, the ending was considered ambiguous.

The exact predictability percentages are shown in Appendix B. In addition to nouns being classified as ambiguous or unambiguous based on their noun ending, 12 of the target nouns were considered "exceptions" in that their gender was opposite of what their ending predicted. The role of phonological[22] ending ambiguity was considered in order to determine whether participants use noun endings during gender activation. According to Schriefers and Jescheniak (1999), phonology should not play a role for NSs, but may very well play a role for NNSs, especially English learners of French who would have to create from scratch a gender slot in their grammatical information store. However, to date, no gender priming task has considered noun ending gender ambiguity. To control

[22] In order to simplify the noun-ending ambiguity variable, only phonological noun endings were included. There were no nouns with morphological endings.

for potential gender activation from other sources, none of the stimuli had semantic gender (e.g., girl), and nouns with morphological endings were avoided.[23]

Finally, to minimize L1-L2 gender congruency effects (i.e., a target noun's gender differs between the L1 and L2), the stimuli do not include Spanish-French or Dutch-French cognates.[24] English-French cognates should not pose a problem because English words do not have gender.

7.3.2 Design

Each target picture was preceded by a gender prime that was either gender congruent, gender incongruent, or gender neutral. The gender congruent and incongruent primes were definite determiners (*le* [the, masc.], *la* [the, fem.]), indefinite determiners (*un* [a, masc.], *une* [a, fem.]), possessive pronouns (*mon* [my, masc.], *ma* [my, fem.]), and subject pronouns, (*il* [he], *elle* [she]). The gender neutral prime was *chaque* (each), which has the same phonologic and orthographic form in both masculine and feminine. For example, in a gender congruent condition, *le* primes the target picture *livre* (book, masc.); in a gender incongruent condition, *le* primes the target picture *chaise* (chair, fem.); and in the gender neutral condition, *chaque* primes the target picture *table* (table, fem.). As discussed in Chapter 3, Alario et al. (2004) included only syntactically compatible prime-target combinations; the current task included gender congruent and gender incongruent *incompatible* prime-target combinations, such as *il* + *livre* (he + book). Including

---

[23] One of the 48 target nouns, *cendrier (ashtray)*, has the morphological ending *–ier*, however, because this masculine ending does not conflict with the masculine phonological ending [ɛ], it is not considered problematic.

[24] Despite excluding cognates, L1 gender activation may still occur. Dutch L1 gender activation should not interfere with French L2 gender activation given that the genders are different (common and neuter in Dutch vs. masculine and feminine in French), but Spanish L1 gender activation may interfere. Of the 48 target nouns, 15 have incongruent Spanish-French gender (i.e., squirrel is masculine in French, but feminine in Spanish). However, within the limits of this study, it is not possible to determine the role of L1-L2 gender congruency.

incompatible prime-target combinations will determine whether previous congruency effects in gender priming tasks are due to determiner-noun pair co-occurrence frequency effects or to activation of gender node.

In addition, nine "catch" trials, in which a prime was followed by a "?", were included. In these trials the participants were asked to report the prime they just saw. According to Alario et al. (2004), this enhances processing of the prime throughout the experiment and prevents the participant from learning to ignore the prime.

The target pictures were divided into three lists of 16 (1-16, 17-32, 33-48), as shown in Table 5, such that participants saw each picture once (total of 48), but across all participants, the pictures appeared in each of the three congruency conditions an equal number of times. Gender, frequency, name agreement, and phonological ending ambiguity were also evenly distributed across lists (see Appendix C for complete list of materials)

After the target words were divided into three lists, each target word was randomly assigned to a prime type, that is, a definite determiner, indefinite determiner, possessive pronoun, or subject pronoun. The type of prime was consistent across groups; for example, if the target picture *drapeau* (flag) was assigned to the prime *le* in the congruent condition, it was assigned to the prime *la* in the incongruent condition. Seven of the target pictures begin with a vowel, therefore, these pictures were not assigned to primes that were phonologically incompatible. For example, *ampoule* (light bulb) was not assigned to *mon/ma* or *le/la* because only *mon* and *l'* would precede a vowel-initial word. Finally, each target picture was assigned to *chaque* in the neutral condition.

Table 5

Gender Priming Task Conditions

Group	Congruent		Incongruent		Neutral	
Group A	List 1 (1-16)		List 2 (17-32)		List 3 (33-48)	
	8 masc.	8 fem.	8 masc.	8 fem.	8 masc.	8 fem.
	le (2)	la (2)	la (2)	le (2)	chaque (8)	chaque (8)
	un (2)	une (2)	une (2)	un (2)		
	mon (2)	ma (2)	ma (2)	mon (2)		
	il (2)	elle (2)	elle (2)	il (2)		
Group B	List 2 (17-32)		List 3 (33-48)		List 1 (1-16)	
	8 masc.	8 fem.	8 masc.	8 fem.	8 masc.	8 fem.
	le (2)	la (2)	la (2)	le (2)	chaque (8)	chaque (8)
	un (2)	une (2)	une (2)	un (2)		
	mon (2)	ma (2)	ma (2)	mon (2)		
	il (2)	elle (2)	elle (2)	il (2)		
Group C	List 3 (33-48)		List 1 (1-16)		List 2 (17-32)	
	8 masc.	8 fem.	8 masc.	8 fem.	8 masc.	8 fem.
	le (2)	la (2)	la (2)	le (2)	chaque (8)	chaque (8)
	un (2)	une (2)	une (2)	un (2)		
	mon (2)	ma (2)	ma (2)	mon (2)		
	il (2)	elle (2)	elle (2)	il (2)		
Catch trials Prime + "?"	One for each prime (total of 9)					

7.3.3 Procedure

Participants were randomly assigned to one of the three counterbalancing groups (A, B, C; see Table 5) and tested individually. Prime words and target pictures were presented in random order on the screen of a 14-inch laptop computer equipped with Psyscope (Cohen, MacWhinney, Flatt, & Provost, 1993). For each trial, participants saw the following sequence: a fixation cross appeared for 500ms, followed by a prime word in size 24 font for 250ms. The screen was then clear for 64ms before the target picture appeared. The target picture remained on the screen until the participant responded or until it timed out after 4000ms. The screen was clear for 2500ms before the next trial began.[25]

Participants were told that they would see a word followed by a picture, and their task was to name the picture in French as quickly and accurately as possible. From time to time, a question mark (instead of a picture) would follow the word, and in this case they were to repeat the word they had just seen as quickly and accurately as possible. The participant's RTs were measured by a voice activated microphone and button box and the responses were recorded by a digital recorder. Participants were shown two example trials, ten practice trials, and given the opportunity to ask questions before the start of the experiment. All instructions were presented in French.

As discussed above, previous studies show a gender priming effect when a lemma of a gender-marked determiner activates its grammatical gender node, which in turn, activates all nouns of that gender, subsequently facilitating production of a gender congruent target noun, or inhibiting production of a gender incongruent target noun. That is, a gender congruent prime will boost activation of the target noun, resulting in faster naming times as compared to the neutral condition; a gender incongruent prime will boost activation of target nouns of a different gender than the target noun, thus, resulting in slower naming times as compared to the neutral condition. If NNSs do not show a

[25] This procedure is identical to that of Alario et al. (2004), except for the timing out of the target noun after 4000ms. In Alario et al., the target picture remained on the screen until the participants responded, however, because this study included NNSs, imposing a time limit was intended to prevent a participant from spending long periods of time attempting to retrieve the name of a noun they might not know.

gender priming effect, it indicates that they do not have access to a gender storage and nodal system.

The following hypotheses elaborate on the first general hypothesis presented above:

1a. Spanish learners of French will show evidence of a gender storage and nodal system, as revealed by faster RTs in the congruent condition than the incongruent condition.

1b. Dutch learners of French will show evidence of a gender storage and nodal system, as revealed by faster RTs in the congruent condition than the incongruent condition.

1c. English learners of French will not show evidence of a gender storage and nodal system, as revealed by similar RTs in the congruent and incongruent conditions.

7.3.4 Pilot

To ensure that this task was functioning properly, that is, producing congruency effects for NSs similar to those found in Alario et al. (2004), it was piloted with a total of eight French NSs. The first five NSs who participated grew up in a French-speaking home in France and began learning English at school between 11 and 15 years of age. All moved to the U.S. in adulthood (after age 20) and their mean number of years in the U.S. (or another English speaking country) was 10 years (range 4-21 years). All reported using primarily English in the community and at home, although one reported using both French and English at home, two reported using French at work, and one reported using French within the French community. Four of the participants return to France once a year for 1-4 weeks, and one participant returns to France for two weeks every two years.

First, the participants' voice responses were coded for accuracy. For the analysis, only the target trials were considered; catch trials (prime + "?") were not included. Trials in which the wrong response was provided or the correct response was preceded by "uh" or any other type of stutter, were removed from the data. Next, the mean RT for each condition (individually for each participant) was calculated,[26] and any trials for which the RT was 2 standard deviations above or below the mean were removed from the data. A total of 32 trials were removed (of 240, 13.3%).

The mean RT for each condition for each NS was calculated (Table 6). Because the purpose of this pilot was to determine whether the trends in the data replicate Alario et al. (2004), their results are presented in Table 7 below.

Table 6

Mean RTs (ms) for 5 NS Pilot Participants

Subject	Congruent	Neutral	Incongruent
1	1000	968	1070
2	893	870	886
3	829	903	886
4	1094	1015	1030
5	1132	1103	1063
Mean	990	972	987

[26] By removing outliers from all three conditions combined, more RTs may be removed from one condition. Because we expect the RTs to be different for each condition, removing outliers per condition will prevent favoring one condition over another.

Table 7

Alario et al. (2004, p. 196) Results

Picture Targets	Mean RT (std dev)
Congruent	703 (53)
Non gender-marked (chaque)	725 (63)
Incongruent	754 (101)
Congruency effect	+ 22
Incongruency effect	- 29

The overall means of the five pilot NSs did not replicate Alario et al.'s (2004) findings: RTs in the neutral condition were the fastest, and the RTs in the congruent condition were slower than the incongruent condition.

One possible explanation for the lack of congruency effects is the presence of the *il/elle* primes, which were not included in Alario et al. (2004). These primes differ from the others (*le/la*, *mon/ma*, *un/une*) in that they are not syntactically compatible with the pictures. For example, "*la* + *chaise*" is a possible NP, whereas "*elle* + *chaise*" is not. That is, it may be that the *il* and *elle* primes do not facilitate or interfere with picture naming as determiners do and removing those items from the analysis would uncover the congruency effects. Therefore, the mean RTs for each condition for each NS were calculated a second time, excluding all trials with the *il* and *elle* primes. The results are shown in Table 8.

Table 8

Mean RTs (ms) Without Il and Elle Primes

Subject	Congruent	Neutral	Incongruent
1	909	968	1091
2	928	870	918
3	813	903	913
4	1063	1015	1005
5	1057	1103	1097
Mean	954	972	1005

The trend matches that of Alario et al.'s (2004) findings. The RTs in the congruent condition are fastest, and the RTs in the incongruent condition are slowest. However, the overall RTs are slower than those found in Alario et al. A possible explanation for this difference is the language mode of the participants. Alario et al.'s participants were French NS university students, all living in France at the time of the experiment. The NSs in the current pilot have been living in the U.S. for 4-21 years and all use English in their daily lives. It may be that being immersed in a non-L1 environment affected the NSs' ability to name the target pictures.

To explore this possibility further, three additional NSs completed the task. These three NSs had been in the U.S. for less than a year and used French in their job and both French and English at home and in the community. The results of all eight NSs combined, including all prime types (presented in Table 9) show both facilitation and interference effects, replicating Alario et al. (2004), though the effects are minimal (9ms and 7ms respectively). The overall RTs are slightly faster with the additional three NSs, although still slower than those found by Alario et al.

Table 9

Mean RTs (ms) for 8 NS Pilot Participants

Subject	Congruent	Neutral	Incongruent
1	1000	968	1070
2	893	870	886
3	829	903	886
4	1094	1015	1030
5	1132	1103	1063
6	842	834	853
7	960	1055	1087
8	818	894	821
Mean	946	955	962

Finally, RTs without the *il/elle* prime trials are considered. The results are similar to those found by Alario et al. (2004), as shown in Table 10, and the difference between the congruent and neutral conditions is robust. That is, both facilitation and interference congruency effects are evident.

Table 10

Mean RTs (ms) Without Il and Elle Primes

Subject	Congruent	Neutral	Incongruent
1	909	968	1091
2	928	870	918
3	813	903	913
4	1063	1015	1005
5	1057	1103	1097
6	828	834	851
7	932	1055	1046
8	814	894	834
Mean	918	955	969

The results of the NS pilot study indicate that gender congruency effects may be masked by slower RTs due to participants living in an L2 environment, as well as by the presence of *il/elle* primes. Therefore, only French NSs who were currently living in France were eligible to participate as the NS controls. Because removing the *il/elle* prime trials from the analysis unmasked the congruency effects that occurred with syntactically compatible primes, that is, the presence of the syntactically incompatible primes did not prevent congruency effects from occurring, these primes remained in the task for the main study. An analysis of the RTs for these primes as compared to the syntactically compatible primes was planned for the main experiment.

Based on the NS pilot results, a power analysis was conducted to determine the appropriate number of participants (both NS-control and Dutch, English, and Spanish learners of French) needed to complete this task in the main experiment. Using the Java Applets for Power and Sample Size (Lenth, 2006), the NS pilot data were entered into a one-sample t-test (one-tailed, alpha .05), with the desired power set at 0.8, which is considered as a standard for acceptable power. Results indicated that a sample size of 20 NSs was necessary. Because NNS gender priming effects were likely to be smaller and/or have more variability, the NNS sample size was doubled. To this end, approximately 20 NS controls and 40 NNSs from each language group were recruited to participate in the main experiment.

After the pilot study with NSs was completed, the task was piloted with six NNSs (L1 English) to determine whether a familiarization task was necessary. Whereas NSs were able to accurately name the target pictures, it was uncertain whether this would be the case for NNSs. The six NNSs were those advanced learners who piloted the pre-screening task. Their target picture naming accuracy ranged between 73-98% (mean 87%), indicating their ability to accurately name the target pictures without a familiarization task. In addition, although NNS accuracy was not as high as NS accuracy, and might be improved with the addition of a familiarization task, such a task would introduce additional variables that would interfere with the main experiment. Specifically, the purpose of a familiarization task is to make known to the participants the intended name of the target picture, for example, *orteil* (toe) as opposed to *pied* (foot), or *manteau* (coat) as opposed to *veste* (jacket). However, if a participant did not know the name of the target picture, for example, *robinet* (faucet), and was introduced to the word in the familiarization task, this could introduce a memory component during the main experiment that would mask potential gender priming effects. That is, the participant would be searching for the word presented in the familiarization task rather than responding directly to the picture. Furthermore, the purpose of the priming task was to determine whether there are gender congruency effects for words *known* by the participants; blank cells for unknown words would be preferred over RTs that reflect a memory search for new words presented in the familiarization task, especially considering there is no way to determine which words were unknown prior to the task.

7.4 Grammaticality Judgment Task

A grammaticality judgment task using a rapid serial visual presentation (RSVP) paradigm was administered to investigate the participants' gender agreement accuracy during processing.

7.4.1 Materials

Forty-eight target sentences contained noun-adjective gender agreement errors. Half of the sentences contained feminine target nouns and half contained masculine target nouns. The adjectives modifying the feminine nouns were masculine and occurred either close to (directly following) the noun or far from (at least 4 words after) the noun; conversely, the adjectives modifying the masculine nouns were feminine, with the same close and far conditions. To avoid end-of-sentence effects, the adjective containing the error never occurred in the sentence final position. In addition, half of the sentences contained determiners providing a gender cue for the target noun, always occurring directly before the target noun; the other half did not contain gender cues.

The target nouns were selected from the www.lexique.org database (New et al., 2001) and were controlled across conditions for word frequency, length, and number of syllables (see Table E.1 in Appendix E for details). As with the gender priming task, target noun phonological endings were either ambiguous or unambiguous, based on the same criteria as used in the gender priming task. The exact predictability percentages are shown in Appendix D. While phonological ambiguity was not an independent variable, ambiguous and unambiguous endings were evenly distributed among conditions so as not to provide additional gender cues for one condition over another.[27] Semantic groups that share gender regardless of the phonological ending (e.g., days of the week), nouns with semantic gender (e.g., girl, boy), and nouns derived from verbs (which tend to be feminine) were excluded. Furthermore, an effort was made to avoid noun-adjective collocations in the close condition. Errors in typical noun-adjective combinations, such as *cauchemar effrayante* (frightening nightmare), may be easier to detect because the correct version of this combination is likely to be more frequent than a less typical combination, such as *cauchemar affolante* (terrifying nightmare). Because phonological form is activated during reading, only adjectives with phonologically distinct masculine and feminine forms were included to ensure that the ungrammatical words in the target sentences all received the same degree of activation. Finally, all adjectives were used only once except for *blanc/blanche* (white), which was used twice.

In addition to the 48 target sentences, 96 filler sentences were included. Twenty-four filler sentences contained errors similar to those included in the pre-screening task (complex negation, *avoir* vs. *être* verb form, and subject-verb agreement), allowing for an additional proficiency measure. An effort was made to include only errors that are phonologically realized; however, two of the twenty-four incorrect fillers have errors that are orthographically, but not phonologically, realized. For example, **ils arrive* (they arrive) is phonologically identical to its correct form, *ils arrivent*. The remaining twenty-two sentences contain errors that are both orthographically and phonologically realized.

[27] It was stated in the proposal for this project that only ambiguous endings (less than 70% predictability by both Surridge (1993, 1995) and Lyster (2006) would be included; however, this constraint was too limiting in terms of selecting target nouns and creating sentences that met the other constraints.

In order to have an equal number of correct and incorrect sentences, the remaining 72 filler sentences contained no errors and were generally similar in terms of length and structure to the target and incorrect fillers.

All sentences were carefully reviewed by a professor of French. After revisions were made, a French NS who was unfamiliar with the project (and, therefore, unfamiliar with the stimuli constraints and task conditions) reviewed the sentences in order to ensure that the sentences were native-like and there were no errors other than the ones intended. The NS suggested several revisions. Because two of the suggested changes did not respect the constraints (i.e., adjective must occur at least four words after the target noun in the far condition), the researcher and NS worked together to revise the sentences until they were correct and appropriate for the condition. An example of each sentence type is presented in Table 11; the complete set of sentences is provided in Appendix E.

Table 11

Grammaticality Judgment Task Example Sentences

Condition	Sentence Example
Close with cue	* La boîte lourd qui se trouve dans le grenier appartient à ma mère. (The ancient box that is in the attic belongs to my mother.)
Close, no cue	* Le jeune étudiant n'a pas fait de cauchemar affolante depuis son enfance. (The young student hasn't had terrifying nightmares since his childhood.)
Far, with cue	* Le marin insiste pour que la voile de son bateau soit léger, malgré le prix. (The sailor insists that the sail on his new boat be light, despite the price.)
Far, no cue	* Leur fierté d'avoir gagné ce match était bien apparent sur leurs visages. (The pride of having won the match was obvious on their faces.)
Filler – negation	* Il ne peut continuer plus ses études car il doit travailler à plein temps. (He can no longer continue his studies because he must work full time.)
Filler – verb form	* Ma grand-mère m'est offert une armoire l'année dernière. (My grandmother offered me a cupboard last year.)
Filler – subject-verb agreement	*Les enfants sont choisi les mêmes jeux chaque jour pendant tout l'été. (The children chose the same games every day during the summer.)

A one-way ANOVA showed no significant difference ($p > .05$) in sentence length across the four target conditions (close with cue, close with no cue, far with cue, far with no cue, filler incorrect, filler correct), nor was there a significant difference ($p > .05$) in target noun frequency across the four target conditions (close with cue, close with no cue, far with cue, far with no cue).

7.4.2 Design

The gender cue and noun-adjective distance conditions were crossed, as shown in Table 12.

Table 12

Grammaticality Judgment Task Variables

Distance Condition	Gender Cue (Determiner)	No Gender Cue
Adjective 'close' to noun	6 masculine target nouns 6 feminine target nouns	6 masculine target nouns 6 feminine target nouns
Adjective 'far' from noun	6 masculine target nouns 6 feminine target nouns	6 masculine target nouns 6 feminine target nouns

The 48 target sentences and the 96 filler sentences were automatically randomized by the computer program, Psyscope (Cohen et al., 1993), such that each participant saw the 144 sentences in a different random order.

7.4.3 Procedure

In an earlier version of this task[28] the stimuli were presented aurally to the participants, who were asked to determine the grammaticality by push of a button at the end of each sentence. However, there are three main disadvantages associated with that design. First, comprehension difficulties may arise with an aural task. Determiner cues and masculine-feminine adjective distinctions may be phonologically similar ([lə] vs.

[28] The earlier version refers to the grammaticality judgment task used for the qualifying paper that preceded this project.

[la], [vif] vs. [viv]) and, therefore, not salient to the participant, especially in a pre-recorded listening task. Second, there may be something unnatural about the way the NS reads ungrammatical sentences that affects the participants' responses. Third, as mentioned above, it has been shown that requiring participants to wait until the end of a sentence to make a grammaticality judgment adds an additional memory component when the error occurs in the middle of the sentence, as opposed to at the end of the sentence (Sabourin & Stowe 2004).

To eliminate these drawbacks, an RSVP paradigm was used instead. Presenting the sentences visually ensures that the participants are able to notice both the determiner cue and the adjective agreement errors. Visual presentation also eliminates additional variables associated with the NS's production of the sentence (i.e., natural sounding, regional accent). Although one potential drawback to visual presentation is the possibility of reading effects, such as the opportunity to scan back in the sentence to check or confirm agreement errors, the RSVP paradigm presents only one word at a time, with each subsequent word replacing the previous one. There is no opportunity to scan backwards and reread parts of the sentence. In addition, the timing of word presentation is 400ms, which allows for normal reading pace and, therefore, the phonological activation that normally occurs in both normal reading pace and normal auditory presentation (for a discussion of the use of the RSVP paradigm in grammaticality judgment tasks, see Blackwell, Bates, & Fisher, 1996; Rummer, 2004). Furthermore, the automatically paced presentation of the sentences ensures that all participants are exposed to the target error for the same amount of time, as they would be in a listening task, whereas in a self-paced paradigm it would be impossible to control how long each participant spent on each word, resulting in an additional variable. Finally, participants are able to make their grammaticality judgment as soon as they detect an error. They do not need to wait until the end of the sentence, thus, eliminating an additional memory component.

Participants were tested individually. The sentences were presented one word at a time in the center of a 14-inch laptop screen in size 18 font. For each trial, participants saw the following sequence: a blank screen for 3000ms, followed by a fixation cross for 1000ms. The screen was then clear for 500ms before the first word of the sentence appeared. Each sentence appeared one word at a time in the center of the screen, each word appearing for 400ms. At the end of the sentence, the screen remained blank for 3000ms. Participants were instructed to press the "incorrect" button on the button box as soon as they detected an error, even if the sentence was still running, or to push the "correct" button if no error had been detected after the sentence was completed. The sentence continued to run to the end even after an "incorrect" button press. The buttons were labeled "correct" or "incorrect" with an overlay. Participants were instructed to focus on what they considered to be proper grammar, and not on ideal style, punctuation or spelling, which would always be correct.[29] The button box measured the RT and logged the response. All instructions were presented in French. Participants were given six practice trials and the opportunity to ask questions before the start of the experiment.

If participants achieve native-like accuracy on all four target conditions (close with cue, close with no cue, far with cue, far with no cue), this suggests that native-like gender processing procedures are in place. If participants achieve higher accuracy on grammaticality judgments when a gender cue is provided than when no gender cue is

[29] This wording was borrowed from Blackwell et al. (1996) and presented in the task instructions.

provided, this suggests that participants rely on external cues rather than inherently stored gender information. On the other hand, if participants show no difference in accuracy on grammaticality judgments whether or not a gender cue is provided, this suggests they use grammatical gender information inherently stored during gender processing. Furthermore, if participants achieve higher accuracy on grammaticality judgments when the adjective is close to the noun as compared to when the adjective is far from the noun, this indicates a WM component. That is, participants rely on WM to (a) keep the noun's gender activated, or accessible, and (b) remember to carry out noun-adjective gender agreement in order to apply the correct form of the adjective(s) throughout the sentence. The role of WM can be confirmed by examining differences in accuracy between the close and far conditions for participants with a low WM span as compared to participants with a high WM span (as measured by an Operation Span task, described in Section 7.5). If only the low span participants achieve lower scores on the far condition than the close condition, this confirms that WM is playing a role in carrying out gender agreement. This conclusion may be strengthened by a correlation analysis between WM span and accuracy on the close and far grammaticality judgment task conditions.

The following hypotheses elaborate the general hypothesis 2-4 presented above:

2a. Spanish learners of French will achieve near-native accuracy on gender agreement.

2b. Dutch learners of French will not achieve near-native accuracy on gender agreement.

2c. English learners of French will not achieve near-native accuracy on gender agreement.

3a. The availability of external gender cues will not facilitate gender agreement accuracy for Spanish learners of French.

3b. The availability of external gender cues will facilitate gender agreement accuracy for Dutch learners of French.

3c. The availability of external gender cues will facilitate gender agreement accuracy the most for English learners of French.

4a. WM span will not be correlated with gender agreement accuracy for Spanish learners of French.

4b. WM span will be correlated with gender agreement accuracy for Dutch learners of French.

4c. WM span will be correlated with gender agreement accuracy for English learners of French.

7.4.4 Pilot

This task was piloted with five French NSs to determine the appropriate RSVP timing for the NS control group and to revise any sentences on which the NSs' judgments did not match those of the researcher. The five NSs were the same as the first five who piloted the gender priming task. After each NS completed the task, the NS and researcher reviewed all sentences for which the NS's judgment did not match that of the researcher's. The NSs were asked to indicate whether they thought the sentences were grammatical or ungrammatical and provide suggestions for revisions on the ungrammatical sentences.

Of the 48 target sentences, two of the five NSs correctly judged all 48 target sentences. The remaining three NSs judged at least one of the target sentences as correct during the experiment, that is, they did not detect the gender agreement error. Two of these three NSs each judged only one of the 48 target sentences as correct, but identified the gender agreement error when asked to review the sentence a second time, and one judged 6 of the 48 target sentences as correct, but identified the gender agreement error in 5 of the 6 sentences when asked to review the sentences a second time. The researcher asked this NS directly if there was a gender agreement error in the sixth sentence and the NS decided there was no error and the sentence was correct. However, the other four NSs found this same sentence to be incorrect during the experiment.

Of the eight missed gender agreement errors, five were on sentences in the far condition (four in the 'cue' condition and one in the 'no cue' condition) and two were in the 'close no cue' condition. Only one sentence was incorrectly judged by two NSs, the other seven were each incorrectly judged by only one NS. Based on these results, no revisions were made to the 48 target sentences.

Turning to the 96 filler sentences, the NSs judged between three and eight of the sentences differently than the researcher. All five NSs identified the same sentence as needing revision and four of the five NSs identified a second sentence as needing revision. In addition, several minor revisions were suggested. Finally, there were, on average, 3.4 unintentional incorrect button pushes per participant (range 0-6; 2%). The average NS accuracy on this task was 99%, indicating that the timing of the word presentation is appropriate for NSs. After the revisions suggested by the NSs were made (i.e., word choice, incorrect preposition), three additional NSs reviewed the sentences and all agreed with the intended judgments.

The task was also piloted with five NNSs (L1 English) to determine whether the same RSVP timing would be appropriate for NNSs, whether the fillers were functioning as an appropriate proficiency measure, and whether there was a high false alarm rate for the correct sentences, that is, whether NNSs were judging the correct sentences as incorrect for a reason unanticipated by the researcher. The five NNSs were those advanced learners who piloted the pre-screening task.[30] After completing the task, they were asked to assess the difficulty of the speed of stimulus presentation. The NNS scores are shown in Table 13.

Table 13

NNS Pilot Participant Grammaticality Judgment Task Accuracy Scores (%)

	Total Task Score	Target Sentences	Filler Sentences		
			Total	Incorrect Fillers	Correct Fillers
NNS 1	62%	42%	82%	77%	94%
NNS 2	68%	46%	90%	79%	94%
NNS 3	85.5%	79%	92%	92%	92%
NNS 4	92.5%	90%	95%	96%	94%
NNS 5	90%	90%	90%	75%	94%
Mean	79.6%	69.4%	89.8%	83.8%	93.6%

[30] One of the six NNSs who piloted the pre-screening and gender priming tasks did not complete the grammaticality judgment task.

The mean total score for all 144 sentences was 79.6% (range 62-92.5%). The mean score on the target sentences was 69.4% (range 42%-90%). Two participants scored surprisingly low, suggesting that they were unable to detect the noun-adjective gender agreement errors. However, these participants performed well (above 80%) on the filler sentences, indicating that the low target sentence accuracy was not due to low proficiency. Furthermore, these low scores are similar to those found in an earlier version of this study, and are most likely an accurate representation of even advanced NNS performance.

The mean score for the filler sentences was 89.9% (range 82-95%). A closer look at the breakdown between correct and incorrect fillers shows that the incorrect fillers were more difficult, with a mean score of 83.8% (75-96%), as compared to 93.6% (92-94%) on the correct fillers. These scores indicate (a) there was not a high false alarm rate for the correct sentences, and (b) the participants were able to detect complex errors on negation, *avoir* vs. *être* verb form, and subject-verb agreement, indicating their advanced proficiency.[31] Finally, all five NNS pilot participants indicated that the speed of the sentence presentation was appropriate and that the task was difficult, but slowing it down would not have made it easier.

To confirm that the NNSs' accuracy on the filler sentences is indeed indicative of high proficiency, and that they can be used as an additional proficiency measure, three NNSs who had not achieved a high level of proficiency also completed the

[31] The two participants who scored 77% and 79% on the incorrect fillers had not used French on a regular basis in the three months prior to completing this task, and mentioned that they thought they would have done better if they had been in "French mode"; whereas the other three participants had either recently spent time in France or use French on a regular basis for their job.

grammaticality judgment task.[32] The total mean score of these three participants on the filler sentences was 77.8% (range 69-86%). On the incorrect fillers, the low proficiency group scored a mean accuracy of 66.7% (range 50-83%) and on the correct fillers, 81.5% (range 69-88%). The pattern of these accuracy scores is similar to that of the high proficiency group in that the incorrect fillers are more difficult than the correct fillers and there is not a high false alarm rate. The high and low proficiency mean scores are presented in Table 14 for ease of comparison.

Table 14

Grammaticality Judgment Task Filler Accuracy (%) and Range for Low and High Proficiency NNS Pilot Participants

	Total Filler	Incorrect Filler	Correct Filler
Low Proficiency (n = 3)	77.8% (69-86%)	66.7% (50-83%)	81.5% (69-88%)
High Proficiency (n = 5)	91.2% (82-95%)	83.8% (75-96%)	93.7% (92-94%)

The difference in performance between the low and high proficiency groups demonstrates that the grammaticality judgment task filler sentences are more difficult for the low proficiency group, and that the these sentences may serve as an additional proficiency measure.

Based on the NS and NNS pilot results, a power analysis was conducted to determine the appropriate number of participants (NS controls and NNSs) needed to complete this task in order to detect NS-NNS differences. As with the gender priming power analysis,

[32] These three participants are those in the "low proficiency" group who piloted the pre-screening sentences. That is, they are advanced learners of French, but not highly proficient.

the Java Applets for Power and Sample Size software (Lenth, 2006) was used. First, the NS and NNS pilot data on the target sentences were entered into a two-sample t test (one-tailed, alpha .05). Results indicate that with a sample size of 20 (for each group), the power would be .9991.

Next, to determine the appropriate number of NNS participants necessary to detect differences within the target sentence conditions, the English NNS pilot data on the two main manipulations (cue vs. no cue and close vs. far) were considered. Accuracy data (shown in Table 15) were entered into two one-sample t-tests (one-tailed, alpha .05). Results indicate that for the cue vs. no cue manipulation, a sample size of 40 would yield a power of .7303. However, for the close vs. far manipulation, a sample size of 40 would yield a power of .2784. Increasing the sample size to 60 results in a power of .5226, which indicates that based on the pilot data, there may not be a meaningful difference in accuracy between the close and far conditions, regardless of the number of participants.[33]

[33] Although the power analysis indicates that there may not be a meaningful difference in accuracy between the close and far conditions, the pilot was only conducted with English NSs; a difference may be found with one of the other NNS language groups.

Table 15

NNS Pilot Participant Accuracy (%) on Cue, No Cue, Close, and Far Grammaticality Judgment Task Target Sentences

Participant	Cue	No cue	Close	Far
NNS1	52%	31%	55%	21%
NNS2	50%	42%	50%	42%
NNS3	75%	83%	71%	83%
NNS4	88%	92%	88%	92%
NNS5	92%	88%	88%	92%

Based on the first two power analyses, and consistent with the gender priming power analysis, a sample size of approximately 40 participants from each of the three NNS language groups was deemed appropriate.

7.5 Operation Span

A measure of WM capacity was included to investigate the role of WM in gender agreement processing. Participants completed an operation span (O-Span) task (Turner & Engle, 1989) in which they were presented with a series of mathematical operations in sets ranging from 2-6, with each expression followed by an L1 word. The participants were asked to indicate whether the expressions were correct and to maintain the sets of words in memory.

This task is based on Turner and Engle's (1989) operations-word task and is used to measure WM capacity; that is, it measures the participants' ability to temporarily store information while simultaneously performing a cognitive task. Both O-Span and reading span tasks (Daneman & Carpenter, 1980) share the same underlying structure in that they each measure the ability to store information while simultaneously performing a cognitive task, and both are considered reliable and valid measures of WM capacity (Conway, et al., 2005); however, an O-Span task, rather than a reading span task, was determined more appropriate for this study for two reasons. First, because participants' WM span scores will be correlated to accuracy on the grammaticality judgment task, which is a language task, and more specifically, a reading task, WM capacity should not be measured through a similar task. In other words, to ensure that any potential correlation between WM span and grammaticality judgment task accuracy is not driven by an underlying factor common to both tasks, a WM task that relies on mathematical operations as the processing component, rather than reading as the processing component, was used. Second, because the WM task was administered in the participants' L1, the O-Span task allowed for consistency in both the mathematical equations and word recall lists across multiple languages.

The purpose of including the O-Span in this experiment was to determine whether accurate adjective agreement is a function of WM capacity, regardless of whether the participant relies on external gender cues or inherently stored gender information. If the participants' accuracy on the grammaticality judgment task, especially on the condition in which the adjective distance is far from the noun, positively correlates with O-Span scores, this suggests that WM capacity plays a role in the participants' ability to effectively hold in their memory the noun's gender throughout the entire sentence in order to provide the correct form of the adjectives modifying the noun.

7.5.1 Materials

Sixty mathematical expressions and 60 English words were included in this task. The words were translated from English into French, Spanish, and Dutch in order for all participants to complete the task in their L1. The mean number of syllables and mean word length for each language is presented in Table 16.

Table 16

Mean Number of Syllables and Mean Word Length for O-Span Words

	Number of Syllables	Number of Letters
English	1.23	4.65
French	1.37	5.05
Spanish	1.90	4.58
Dutch	1.40	4.83

A one-way ANOVA indicated that the word lists differed significantly as a function of number of syllables, F (3, 236) = 23.802, $p < .01$; Tukey post-hoc comparisons showed that the Spanish word list had a significantly higher mean number of syllables than the other languages ($p < .01$), but the other languages did not differ significantly from each other. The greater mean number of syllables in the Spanish word list is most likely due to Spanish nouns typically containing at least two syllables. After replacing six of the Spanish words with shorter Spanish words, the mean number of syllables was still higher. However, because the number of syllables ranged from 1-2 (whereas the other languages had a range of 1-3 syllables) and the mean word length (number of letters) was the shortest of the four languages, no further adjustments were made. A one-way

ANOVA indicated no significant difference for word length between languages ($p > .05$). See Appendix F for the complete list of the O-Span materials.

7.5.2 Design

The 60 mathematical expressions and 60 words were divided into 15 sets ranging from 2-6 expression-word pairs per set, with three sets of each size. Each set had approximately equal numbers of correct and incorrect expressions. The words were always presented in the participants' L1. That is, English participants saw and were asked to recall English words, French participants saw and were asked to recall French words, etc. The sets were presented in random order to prevent the participants from anticipating the number of words to be remembered.

7.5.3 Procedure

Each trial began with a fixation cross for 1000ms, followed by a mathematical expression (e.g., '(18 / 3) – 4 = 2').[34] The participants were instructed to respond as quickly and as accurately as possible by pushing the "correct" button on the button box if the expression was correct and the "incorrect" button if the expression was incorrect. The buttons were labeled "correct" or "incorrect" with an overlay. The expression remained on the screen until the participant responded or until it timed out after 4000ms, at which point a word in the participant's L1 appeared on the screen in size 24 font for 1250ms. At the end of each set, the word RECALL[35] appeared on the screen prompting the

---

[34] Mathematical symbols that are common to all language groups (English, French, Spanish, and Dutch) were used.

[35] The RECALL prompt was translated into the following for each language: RAPPEL (French), WOORDEN (Dutch), PALABRAS (Spanish)

participants to write down the words they had seen on the screen in the order in which they had appeared. The participants were instructed to push a button when they were ready to begin the next set. RT and accuracy on the mathematical expressions were recorded, and the recall lists scored for accuracy.[36] The participants were given three practice sets (with set sizes of two, four, and three) and given the opportunity to ask questions before the start of the experiment. Because the English version of this task was included in a previous study conducted by the researcher, it was not necessary to conduct a pilot for the current study.

7.6 Gender Assignment Post-test

Participants were presented with the written form of the target nouns that were included in the gender priming and grammaticality judgment tasks and asked to indicate their gender (masculine or feminine). The nouns were presented visually in random order on a computer screen. Participants were instructed to push one of two buttons on the button box to indicate whether the word is masculine or feminine (buttons were labeled “masc.” and “fem.” with an overlay). Participants had four seconds to respond before the next word appeared.

The purpose of this task was to assess the participants’ knowledge of the nouns’ gender in order to determine whether a potential lack of congruency effects in the gender priming task and potential low accuracy scores on the grammaticality judgment task target sentences are due to gender information being incorrectly assigned to the lemma.

---

[36] It is possible to incorporate a feedback mechanism to ensure participants perform above a certain percent correct (e.g., 80%) on the mathematical operations; however, based on a pilot study in which only one out of 18 participants scored below 80% correct, it was determined that losing participants due to low operations performance was not a risk and, therefore, a feedback mechanism was not necessary. Engle, Cantor, and Carullo (1992) found similar results in their study in which only 3 out of 70 participants scored below 85%, indicating that poor performance on mathematical equations is uncommon.

In other words, an incongruent gender priming condition may not be incongruent to the participant if he/she has incorrect gender information attached to the target picture. In addition, sentences may be judged as incorrect in the grammaticality judgment task, not because the participant does not have the gender information stored inherently or because he/she cannot retain the gender information in memory throughout the sentence even despite a potential gender cue, but because the gender assigned to the target noun is incorrect to begin with.

7.7 Language History Questionnaire

Participants completed a language history questionnaire (Appendix G) covering general information (age, sex, handedness, and current use of French), and language history (native language, age of onset of French, and number of years and type of learning experiences of French study, and other languages studied). The questionnaire was presented in French to all participants. A modified version excluding questions relating to study of French was given to the NS controls.

7.8 General Procedure

The participants were tested individually in one 90-minute session. The participants were provided with the consent form before their scheduled session. Upon arrival at their session, the participants had the opportunity to ask questions about the consent form, and were asked to sign it (5 minutes). The participants completed the grammaticality judgment task first (30 minutes), which was followed by a 5-10 minute break. Next, participants completed the gender priming task (10 minutes), followed by the O-Span (15 minutes), and finally the gender assignment task (5 minutes). After completing all tasks, the participants filled out the language history questionnaire (5 minutes) and the appropriate compensation paperwork. Each participant was paid 15 Euros for his/her time. Appendix H provides a detailed write-up of the experimental procedure and logistics.

Chapter 8: Results

This chapter presents the results of the experimental tasks. The language history questionnaire data provide an overview of participant characteristics and their language learning experience. Next, the gender assignment post-test and operation span results are presented; because these two tasks are secondary to the main experimental tasks, overall performance will be reported first, with a more complete analysis incorporated into the analyses of the gender priming and grammaticality judgment tasks.

8.1 Language History Questionnaire

The purpose of the language history questionnaire was to obtain a general profile of the participants (age, gender, handedness), their language learning experience (languages studied, age of onset and number of years studied), and for the NNSs, their exposure to French (age of onset, number of years studied, type of exposure), and French proficiency self-ratings. The variables relating to the NNSs' exposure to French were included in the analyses to determine whether performance on the experimental tasks differed as a function of their French language learning experience.

8.1.1 French Native Speakers

Eleven female and ten male NSs of French (FNSs) currently living in Paris, France participated in this study. All reported French as their first language and the only language spoken in their home. Two participants reported left-handedness. The FNSs' mean age was 30.3 years (range 18-65 years). All 21 participants reported having studied at least two foreign languages, including English (21), German (14), or Spanish (7). The mean age of onset to the first foreign language was 11.2 years (range 5-15 years) and the mean length of study was 8.5 years (range 5-13 years). The mean age of onset to the second foreign language was 13.2 years (range 11-14 years) and the mean length of study was 6.1 years (range 4-10 years). Seven participants reported having studied a third foreign language, including Arabic (1), Latin (2), or Spanish (4); the mean age of onset was 18.4 (range 12-21 years) and the mean length of study was 1.8 years (range .5-3 years).

8.1.2 Spanish Native Speakers

Twenty-six female and 11 male Spanish native speakers (SNSs) living in Paris at the time of data collection participated in this study. Twenty-eight were enrolled in a graduate program in France, 10 of whom also reported working while pursuing their coursework, and 9 were working professionals who used French on the job. All 37 participants reported Spanish as their first language and the language spoken in their home during childhood. Their mean age was 28.7 years (range 20-45 years); one of the 37 participants was left-handed. The mean age of onset to French was 17.4 (range 10-31 years) and the mean length of study was 5.9 (range .6-13 years). Because some of the participants reported the number of years they had studied French based only on formal classroom exposure, and because most participants had been using French (either in the classroom or in daily life) since their age of onset, a second measure was calculated by subtracting the participants' age of onset from their age, thus, providing the number of years the participant has known French. This measure is a more accurate representation of the number of years the participants have been learning French, with a mean of 11.2 years (range 2-23 years). The mean number of years the participants had been living in France at the time of the study was 3.1 (range .2-14 years).

Based on a scale of 1 (no ability) to 10 (excellent ability), participants reported a mean rating of 8.3 (range 6-10) for reading, 6.9 (range 3-10) for writing, 7.5 (range 4-9) for speaking, and 8.7 (range 4-10) for comprehension.

Finally, the participants reported their experience with foreign languages other than French. All 37 participants had studied either English (35) or German (2). The mean age of onset to either English or German was 10.8 years (range 1-27 years) and the mean length of study was 9.1 years (range 1-25 years). Twenty-six participants reported having studied a third foreign language, including Catalan (1), Chinese (2), Galician (1), German (8), Italian (7), Portuguese (5), and Russian (2), with a mean age of onset of 21.5 years (range 7-33 years) and a mean length of study of 2.4 years (range .5-10 years). Thirteen participants reported having studied a fourth foreign language, including English (1), German (2), Greek (1), Italian (2), Indian language of the Amazon (1), Portuguese (2), Romanian (1), Russian (2), and Slovak (1). The mean age of onset to the fourth foreign language was 22.5 years (range 10-31 years) and the mean length of study was 1.6 years (range .3-5 years). Two participants reported having studied either Catalan or Chinese as a fifth foreign language, with a mean age of onset of 23.0 years (range 22-24 years) and a mean length of study of .3 years (range .1-.5 years).

8.1.3 Dutch Native Speakers

Thirty female and 8 male Dutch native speakers (DNSs) participated in this study. Nineteen of the DNSs were living in Paris at the time of data collection; of these 19 participants, 7 were enrolled in a graduate program, 3 of whom also reported working while pursuing their coursework, and 12 were working professionals who used French on the job. The 19 remaining DNSs were living in Belgium at the time of data collection; 6

were enrolled in a French teacher trainer program in Brussels and 13 were graduate students in a French linguistics program in Ghent. Brussels is officially bilingual (Dutch and French) and Dutch is the primary language spoken in Ghent. Therefore, although it is possible that the 6 participants living in Brussels had more exposure to French on a daily basis, the participants from Ghent reported using French regularly in their graduate program. All 38 participants reported Dutch as their first language and the language spoken in their home during childhood. Their mean age was 27.8 years (range 20-61 years). Three participants were left-handed. The mean age of onset to French was 10.6 years (range 8-13 years) and the mean length of study was 10.7 years (range 5-18 years). As with the SNSs, an additional measure was calculated by subtracting the participants' age of onset from their age, thus, providing the number of years the participant has known French. The mean number of years the participants had known French was 17.2 years (range 10-50 years). The mean number of years the participants had spent living in France at the time of data collection was 3.9 years (range .08-38 years).

Based on a scale of 1 (no ability) to 10 (excellent ability), participants reported a mean rating of 8.4 (range 5-10) for reading, 7.4 (range 5-10) for writing, 7.5 (range 4-10) for speaking, and 8.8 (range 4-10) for comprehension.

Finally, the participants reported their experience with foreign languages other than French. All 38 participants had studied a foreign language in addition to French, including English (34), German (3), and Mandarin Chinese (1). The mean age of onset to the second foreign language was 12.4 years (range 8-16 years) and the mean length of study was 6 years (range 2-15 years). All 38 participants had also studied a third foreign language, including English (3), German (28), Italian (1), Latin (1), and Spanish (5). The mean age of onset to the third foreign language was 15 years (range 10-32 years) and the mean length of study was 3.9 years (range 2-8 years). Twenty-three participants reported having studied a fourth foreign language, including Arabic (3), German (4), Greek (1), Italian (1), Latin (2), Limburgish (1), and Spanish (11), with a mean age of onset of 17.8 years (range 12-6 years) and a mean length of study of 3.2 years (range 1-6 years). Seven participants reported having studied a fifth foreign language, including Greek (1), Italian (2), German (1), Spanish (1), Indonesian (1), and Japanese (1), with a mean age of onset of 22.1 years (range 14-39 years) and a mean length of study of 2.3 years (.4-4 years).

8.1.4 English Native Speakers

Thirty-one female and 11 male English[37] native speakers (ENSs) living in Paris at the time of data collection participated in this study. Twenty of the ENSs were enrolled in a graduate program, 13 of whom also reported working while pursuing their coursework, and 22 were working professionals who used French on the job. All 42 participants reported English as their first language and the language spoken in their home during childhood. Their mean age was 43.6 years (range 20-67 years). One of the 42 participants was left-handed. The mean age of onset to French was 15.3 years (range 9-38 years) and the mean length of study was 8.0 (range .5-33 years). The mean number of years the participants had known French was 28.2 (range 4-55 years), and the mean number of years the participants had been living in France at the time of data collection was 15.6 (range .6-41 years).

Participants assessed their French proficiency by rating their reading, writing, speaking, and comprehension skills on a scale of 1 (no ability) to 10 (excellent ability).

[37] English NS participants were from the U.S. and the U.K.

The ENSs reported a mean rating of 8.3 (range 4-10) for reading, 6.6 (range 2-9) for writing, 7.5 (range 4-10) for speaking, and 8.8 (range 4-10) for comprehension.

Finally, the participants reported their experience with foreign languages other than French. Thirty of the 42 participants had studied a second foreign language in addition to French, including Dutch (1), Gaelic (1), German (5), Greek (1), Hebrew (1), Indonesian (1), Irish (1), Italian (1), Latin (1), Russian (1), and Spanish (16). The mean age of onset to the second foreign language was 15.57 years (range 2-23 years) and the mean length of study was 3.8 years (range .5-13 years). Nineteen participants reported having studied a third foreign language, including Chinese (1), Dutch (1), Farsi (1), German (6), Greek (1), Italian (2), Latin (2), Russian (2), Spanish (2), and Swedish (1). The mean age of onset to the third foreign language was 21.7 years (range 12-35 years) and the mean length of study was 2.2 years (range .2-10 years). Eleven participants reported having studied a fourth foreign language, including Arabic (1), Finish (1), German (1), Greek (2), Italian (4), Latin (1), and Serbo-Croatian (1), with a mean age of onset of 25.6 years (range 11-66 years) and a mean length of study of 2.9 years (range .1-15 years). Four participants reported having studied a fifth foreign language, including German (1), Hebrew (1), Latin (1), and Spanish (1), with a mean age of onset of 19.3 years (range 12-30 years) and a mean length of study was 1.9 years (range .1-4 years).

8.1.5 Comparison of Non-native Speakers

Table 17 presents the SNS, DNS, and ENS language history questionnaire data for the purpose of comparison. Included are the means (range) for age, age of onset to French (AO), number of years spent in France (Years France), number of years participants had known French (Years Known), self-ratings on reading, writing, speaking and comprehension, and the number of participants who have studied foreign languages other than French.

Table 17

Language History Questionnaire Data for SNSs, DNSs, and ENSs

	SNS n = 37	DNS n = 38	ENS n = 42
Age	28.7 (20-45)	27.8 (20-61)	43.6 (20-67)
AO	17.4 (10-31)	10.6 (8-13)	15.3 (9-38)
Years France	3.1 (.2-14)	3.9 (.08-38)	15.6 (.6-41)
Years Known	11.2 (2-23)	17.2 (10-50)	28.2 (4-55)
Reading	8.3 (6-10)	8.4 (5-10)	8.3 (4-10)
Writing	6.9 (3-10)	7.4 (5-10)	6.6 (2-9)
Speaking	7.5 (4-9)	7.5 (4-10)	7.5 (4-10)
Comprehension	8.7 (4-10)	8.8 (4-10)	8.8 (4-10)
Number of participants who reported studying a foreign language other than French			
Second foreign language	37	38	30
Third foreign language	26	38	19
Fourth foreign language	13	23	11
Fifth foreign language	2	7	4

A series of one-way ANOVAs showed that the three groups differed significantly as a function of age, F (2, 114) = 24.413, $p < .01$, AO, F (2, 114) = 20.039, $p < .01$, years spent in France, F (2, 114) = 22.952, $p < .01$, and number of years participant had known French, F (2, 114) = 22.333, $p < .01$. Tukey post-hoc comparisons for age, years spent in France, and number of years participant had known French showed that all three factors for the ENS group were significantly greater than both the SNS and DNS groups ($p < .01$ for all comparisons), but the SNS and DNS groups did not differ significantly from each other. These differences are consistent with the number of working professionals (22) among the ENSs as compared to the 9 SNS working professionals and 13 DNS working professionals. Based on their experience with French, the ENSs are more "acquirers" and the SNSs and DNSs are more "learners"; however, the majority of ENSs reported studying French in high school (31) and/or college (25) before moving to France, and only six ENSs began learning French after they had moved to France, all of whom also reported having taken French courses in France. Therefore, despite an overall longer French immersion experience, the ENSs have comparable classroom experience to the SNS and DNS participants.

Tukey post-hoc comparisons for AO showed that the DNS group was significantly lower than both the SNS and ENS groups ($p < .01$ for both comparisons), but ENS and SNS groups did not differ from each other. The DNSs showed an earlier and more uniform range of AO than the SNS and ENS participants. This difference is a function of the Belgian foreign language curriculum, and it would be difficult to find DNS participants outside this range, or to restrict SNS and ENS subject recruitment to this range.

Despite these differences, the SNS, DNS, and ENS groups' self-ratings did not differ significantly from each other, and all are likely to have achieved a high level of proficiency given their current use of French in a graduate program or as working professionals who use French on a regular basis. Although independent separate proficiency measure was not administered, it is possible to compare performance on the three language tasks completed for the experiment. Table 18 provides the mean accuracy scores for the three NNS language groups on the grammaticality judgment filler sentences, gender priming picture naming, and gender assignment post-test. While complete data, analyses, and discussion for each of these tasks will be presented in subsequent sections, the purpose of Table 18 is to provide an overview of NNS performance as a means of comparison.

Table 18

NNS Mean Accuracy (% and range) on the Grammaticality Judgment Task, Gender Priming Task, and Gender Assignment Post-test

	SNS	DNS	ENS
Grammaticality judgment filler sentences	86% (74-93%)	89% (77-99%)	86% (58-98%)
Gender priming picture naming	62% (42-85%)	64% (42-94%)	74% (48%-98%)
Gender assignment post-test	90% (74-100%)	91% (69-100%)	89% (70-97%)

A series of one-way ANOVAs showed that the three groups did not differ significantly from each other as a function of grammaticality-judgment filler-sentence accuracy or gender-assignment post-test accuracy; however, the three groups did differ significantly as a function of gender priming picture naming accuracy, $F(2, 114) = 14.658$, $p < .01$. Tukey post-hoc comparisons showed that the ENS group had significantly higher accuracy scores than both the SNS and DNS groups ($p < .01$ for both comparisons), but the SNS and DNS groups did not differ significantly from each other. The ENSs' higher picture naming accuracy indicates a larger vocabulary, which may be a function of the number of years they had spent in France as compared to the SNS and DNS participants.

Overall, all three NNS groups achieved high scores on the grammaticality judgment filler sentences and gender assignment post-test. The gender priming picture naming scores are low; however, it should be taken into account that synonyms and plausible picture names that were not an exact match to the intended picture name were scored as incorrect. Therefore, these low scores are due, in part, to a conservative coding process (see Section 8.4 for a complete description of the gender priming task coding process) rather than a vocabulary deficit.

Finally, all three NNS language groups demonstrate extensive experience with foreign languages other than French. That is, the NNSs are highly multilingual, with all of the SNS and DNS participants, and 30 of the 42 ENS participants having studied a second foreign language, and all the DNS and roughly half the SNS and ENS participants having studied at least a third foreign language. The impact of multilingualism on the experimental tasks will be addressed in Chapter 9.

To conclude, the SNS, DNS, and ENS participants demonstrated similar French language experience and comparable French proficiency, that is, no group stands out as being better or worse off than the other groups across all factors. Analyses and results for the experimental tasks will be presented separately for each language group; however, cross-language comparisons will be made at the end of each task section, as well as in the general discussion.

8.2 Gender Assignment Post-test

The data coding process and overall performance for each language group on the gender assignment post-test is presented in this section; however, as the primary purpose of this task was as a follow-up to the gender priming and grammaticality judgment tasks, a more complete analysis of the data will be presented, as relevant, in Sections 8.4 and 8.5.

A flaw in the word list became apparent after the FNS participants completed this task. The feminine noun *voile*, meaning 'sail of a boat', is a masculine noun when the meaning is 'a veil'. Although only 3 of the 21 FNS participants assigned feminine to the intended masculine meaning of *voile*, this trial was removed from the post-test analysis for all FNS participants.[38] On the remaining trials, FNSs scored a mean accuracy of 98% (*SD* = 3%), with incorrect gender assignment on only 39 of the 1953 trials. Following standard procedure, these 39 trials were removed for the RT analysis in order to conduct the analysis on correct trials only. Next, RT outliers that were 2.5 standard deviations above or below the mean were removed, as is common in the literature. To this end, 58 trials (of the remaining 1914; 3%) were removed.

---

[38] This word was changed to "voile (de bateau)" in the gender assignment task for the NNS participants. In addition, one word, "vols" (flights) was presented as a plural; this was changed to the singular form for the NNS participants.

The SNSs scored a mean accuracy of 90% (*SD* = 7%), with 353 inaccurate responses out of a total of 3478. These trials were removed from the RT analysis. An additional 97 trials (of the remaining 3125; 3%) that were 2.5 standard deviations above or below the mean were removed as outliers. The DNSs scored a mean accuracy of 91% (*SD* = 7%); 339 (of 3572) trials with inaccurate responses were removed from the RT analysis. Of the remaining 3233 trials, 98 (3%) were removed as outliers. Finally, the ENSs scored a mean accuracy of 89% (*SD* = 6%), with 449 (of 3948) trials removed due to inaccurate responses. Of the remaining 3499 trials, 112 (3%) were removed as outliers. Mean accuracy (%) and RT (ms) for each of the four language groups are presented in Figures 4 and 5 respectively.

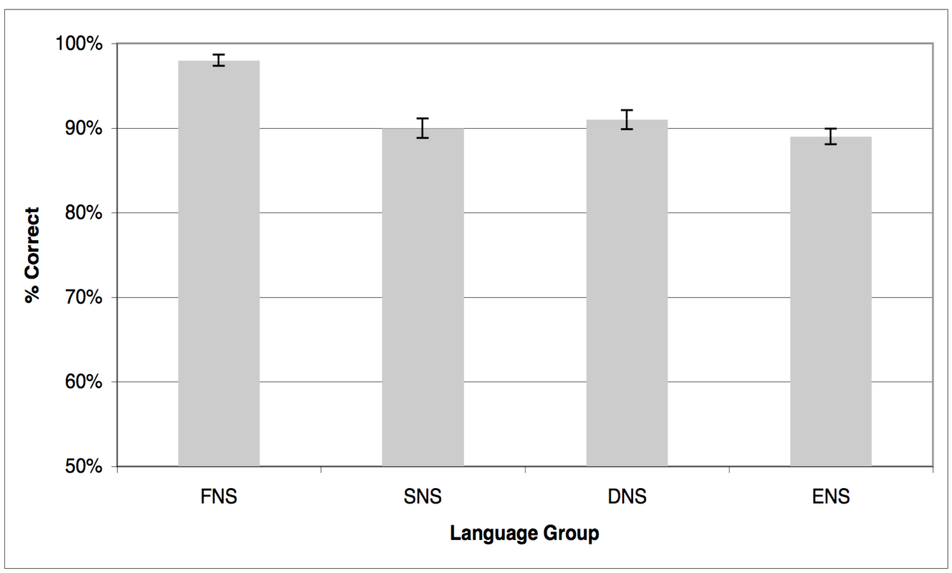


Figure 4. Mean accuracy (%) on the gender assignment task for all four language groups

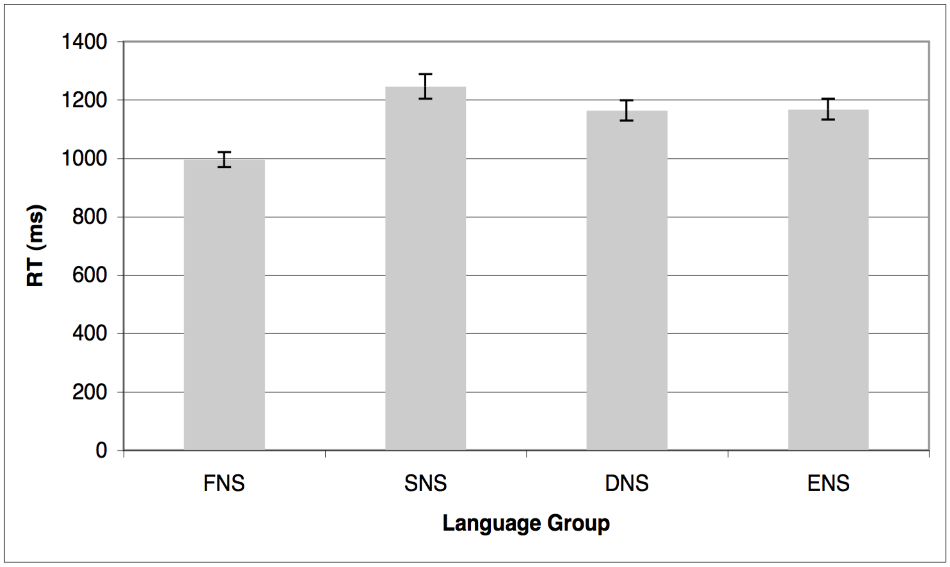


Figure 5. Mean RT (ms) on the gender assignment task for all four language groups

All participants achieved high accuracy on this task, demonstrating their knowledge of French noun gender; however, one-way ANOVAs showed that the four groups differed significantly in both accuracy, F (3, 134) = 11.673, $p < .01$, and RT, F (3, 134) = 5.664, $p < .01$. Tukey post-hoc comparisons showed that the FNS group was significantly more accurate and faster than the SNS, DNS, and ENS groups (accuracy: $p < .01$ for all comparisons; RT: $p < .01$ for the DNS and ENS groups comparisons, $p < .05$ for the SNS group comparison), and the SNS, DNS, and ENS groups did not differ significantly from each other for either accuracy or RT. The breakdown of performance on nouns from the gender priming and grammaticality judgment tasks, as well as an analysis of accuracy based on noun ending ambiguity and gender, will be presented in Sections 8.4 and 8.5 in conjunction with the gender priming and grammaticality judgment analyses.

8.3 Operation Span

The participants' accuracy on the mathematical expressions was scored first. The criterion for excluding participants based on low math accuracy was set at 65%, which was determined after considering math accuracy scores for all four language groups. While a minimum math accuracy score is usually set at 85%, a lower score may be acceptable when the sample size is relatively small (personal communication with Michael Bunting). The math accuracy for the NNS language groups, particularly the SNS group, was lower than expected considering that the task materials (equations and recall lists) were used in a previous version of this study and did not yield low math accuracy scores. One potential explanation is that the task was more demanding for the NNSs as they were immersed in their L2 (French) but completed the O-Span task in their L1. Unfortunately, this could not be avoided, as it was important to recruit participants who use French in their daily lives. However, the participants with low math scores above the cutoff (between 65-75%) also had low recall scores, indicating that there was not a math-recall tradeoff, that is, these participants were allotting attention to the mathematical expressions, as instructed, rather than ignoring them in order to focus solely on remembering the words. Based on the 65% match accuracy cutoff, one FNS, eight SNSs, one DNS, and five ENSs were excluded from further analyses that included O-Span scores.

Next, word recall lists were coded for accuracy. Only words recalled in the correct serial order were accepted. Absolute span was calculated for each participant by counting the total number of words recalled in only the sets that were complete, that is, sets that contained all of the words in the correct serial order. Table 19 presents the mean math accuracy and absolute span scores (out of 60) for each of the four language groups.

Table 19

Math Accuracy (%) and Absolute Span Scores for All Four Language Groups

Language Group	Math Accuracy (range)	Absolute Span (range)
French	88% (77-100%)	32.0 (11-60)
Spanish	83% (67-95%)	28.4 (7-48)
Dutch	91% (72-100%)	30.9 (2-54)
English	86% (65-100%)	26.5 (0-60)

Participants were divided into low and high span groups. In order to create span groups that were consistent across all four language groups, that is, low and high span scores that were based on the same criteria for everyone, all 123 absolute span scores were arranged in order from lowest to highest. The middle 30 scores[39] (range 27-32) were removed, creating two distinct span groups, with scores in the low group ranging from 0-26 and scores in the high group ranging from 33-60. Table 20 presents the number of participants in each language group for the low and high span groups and the mean span scores (and range). The low and high span groups were included in the grammaticality judgment task item analyses as the within-items factor, and individual span scores were included in the correlation analyses.

[39] A common procedure is to include the top and bottom quartile in order to create two distinct span groups; however, this method requires a large number of participants and also would have resulted in an unequal number of participants in the low and high span groups within each language group. Therefore, only the middle 30 span scores were removed, preserving an equal number of low and high span participants for each language group.

Table 20

Number of Participants (n) and Mean (M) Absolute Span Score (Range) for Low and High Span Groups

Language Group	Low Span	Removed	High Span
French	n = 6 M = 20.7 (11-26)	n = 6 (27-31)	n = 8 M = 43.4 (33-60)
Spanish	n = 14 M = 19.6 (7-26)	n = 5 (28-32)	n = 10 M = 40.4 (35-48)
Dutch	n = 9 M = 16.9 (2-25)	n = 12 (27-32)	n = 16 M = 39.3 (33-54)
English	n = 18 M = 13.3 (0-24)	n = 7 (27-32)	n = 12 M = 44.0 (33-60)

Finally, all span scores were entered into a correlation matrix in order to compare WM span to participant age, AO, the number of years spent in France, and the number of years a participant had known French. For the FNS, DNS, and ENS language groups, there were no significant correlations between span scores and the other variables. However, for the SNSs, span correlated significantly with both age, $r(29) = -.636$, $p < .01$, and AO, $r(29) = -.480$, $p < .01$; the negative correlations indicate that SNSs who were older or began learning French later had lower span scores. Although these correlations were not found for the other language groups, it may be that the older SNS participants had more difficulty completing the task. This finding is surprising, however, because the ENSs had a greater mean age as well as a larger age span, but did not show a correlation for age and span score. Regarding the negative correlation between span score and AO, it is possibly a reflection of the SNSs' difficulty in completing the O-Span task in their L1 while immersed in the L2. In other words, the SNSs who began learning French at a later age were more affected by the task of suppressing the L2 in order to compete the recall in the L1. However, this explanation is only speculation. Furthermore, it should be noted that only 29 SNS participants were included in the correlation, and, thus, the results should not be overinterpreted.

8.4 Gender Priming Task

The following data removal process was applied to all four language groups prior to running the analysis. First, accuracy for the catch trial responses was calculated to ensure that they were functioning properly in that the participants were reading the primes throughout the task. Because catch trial accuracy was nearly perfect for all four language groups (98.6%-99.4%), catch trials were not considered in any subsequent analyses. Next, target trials were coded for accuracy. Only the first word produced for each trial was coded, even if a participant corrected him/herself directly after. False start responses, that is, responses that were preceded by a stutter or “euh”, were not accepted, nor were words that were not an exact match to the intended picture name.[40] Therefore, synonyms or other plausible names for a picture were not accepted. The purpose of this conservative coding approach was to ensure that a participant’s accuracy on the gender assignment post-test reflected his/her knowledge of the target pictures’ gender in the priming task. For example, if a participant produced the word “alliance” (wedding ring) for the intended word “bague” (ring), it would be impossible to determine whether that participant knew the gender of “alliance”, as this word was not included in the gender assignment post-test. While the post-test was primarily included to determine the NNSs’

[40] There were no pictures that elicited an incorrect response from a large number of participants; that is, there were no pictures that were especially problematic for any of the four language groups.

knowledge of the target pictures' gender, it was necessary to follow the exact coding process for both the FNSs and NNS learners.[41]

In addition to excluding incorrect responses, trials for which a participant incorrectly assigned gender to a target noun in the gender assignment post-test were excluded. For example, if a participant incorrectly determined the masculine word *chapeau* (hat) to be feminine in the post-test, the trial containing the target picture of a hat in the gender priming task was removed from analysis. Whereas FNS gender assignment errors were likely to be inadvertent button pushes rather than representative of incorrect gender knowledge, this is not the case for NNSs, who may or may not know the gender of a French word. In order to determine whether NNSs show priming effects in the gender priming task, only trials for which a participant knows the gender of the target noun may be included; therefore, all trials for which NNSs incorrectly assigned gender to the target picture in the post-test were removed, making it necessary to follow the same procedure for the FNSs.

RTs were trimmed such that responses faster than 400ms or slower than 2000ms were removed. Responses faster than 400ms were most likely due to the voice activated microphone registering a non-voice signal (e.g., the participant's breath) rather than a voice response. Responses slower than 2000ms were excluded because by the time the participant had retrieved the name of the target picture for production, any automatic activation of the prime's gender would have diminished and a prime-target congruency effect would not occur. Finally, trials with RTs 2.5 standard deviations above or below the mean were considered outliers and removed from the analysis, as common in the

[41] Accuracy scores in the congruent, incongruent and neutral conditions were compared for each language group; data removal based on inaccuracy was evenly distributed across the three conditions for each language group.

literature. As different overall RTs were expected for each of the three conditions, outliers were calculated separately for the congruent, incongruent, and neutral conditions. The number of trials removed based on these criteria will be reported separately for each language group.

Finally, due to the large age range in all four language groups, overall RT on the priming task was correlated to age to determine whether age was related to participants' ability to complete this task. No significant correlation between age and overall RT was found for the FNS, SNS, or ENS participants; however, the DNSs showed a significant negative correlation, $r(38) = -.388$, $p < .05$. The older participants were faster at naming pictures than the younger participants. As the main concern was that the older participants would have difficulty with a task for which the dependent variable was RT, this inverse relationship is not a concern, and most likely represents a higher level of proficiency among the older DNS participants.

Because the nature of the task, which relies on production data, inevitably leads to incomplete data due to the high percentage of inaccurate picture naming responses (especially for NNS participants), multilevel modeling, or hierarchical lineal modeling (HLM) was deemed most appropriate for the analysis. Two main advantages of HLM motivated this decision. First, HLM allows both subject and items to be considered simultaneously in a single regression model, and treats items as random factors, thereby more accurately representing the nature of language items as a subset of all possible items, rather than as fixed factors (see Locker, Hoffman, & Bovaird, 2007 for a complete discussion of the advantages of HLM in psycholinguistic research). Second, and especially critical to the current analysis, HLM does not aggregate data across subjects or items. Instead, raw data for each participant on each trial is included in the data set. According to Locker et al., "…the multilevel model also uses full information maximum likelihood as a means of directly addressing unbalanced or incomplete data, and, thus, complete cases are not required" (p. 724). Employing HLM alleviates the concern of removing a large percentage of data for the NNS participants based on inaccurate picture naming responses.

Because the purpose of this task was to determine whether any of the four language groups demonstrated gender priming effects, and not whether any one group showed "more" or "less" priming than another, separate models were run for each language group. The following steps were followed in running the HLM analysis:

1. An empty (baseline) model with no predictors was prepared.
2. Random subject and item effects were added to the model.
3. Main effects (predictor variables) were added to the model.
4. All 2- and 3-way interaction effects were added to the model.
5. Non-significant interactions were removed in a step-wise fashion from highest order down until the most parsimonious model was achieved.

Before proceeding to each subsequent step, the model was compared to the previous, simpler model, to ensure the added layer of complexity improved the fit to the data. For example, the main effects model was confirmed to have improved the fit to the data over the previous random subject and item effects model. Fit statistics for each model are provided in Appendix I.

In a first analysis, RT was modeled as a function of Counterbalancing Group, Congruency (congruent, incongruent, neutral), Prime Compatibility (syntactically compatible prime, syntactically incompatible prime), Word Ambiguity (ambiguous, unambiguous, exception), and Word Frequency as within-subjects fixed factors. This model was run for all four language groups. However, because language experience factors such as AO, number of years spent learning French (Years Known), and number of years spent in France (Years France), may play a role in the presence or absence of gender priming effects for the NNSs, a second analysis was carried out for the NNS groups to determine the role of these additional factors. The final models containing significant main effects and interactions[42] for each language group will be presented separately, followed by a discussion section comparing the results across all four language groups.

8.4.1 French Native Speakers

Of 1008 trials, 97 (10%) were removed due to false starts or incorrect responses,[43] 20 (2%) were removed due to incorrect gender assignment in the post-test, 4 (.4%) were removed due to RTs faster than 400ms or slower than 2000ms, and 19 (2%) were removed as outliers. Overall, a total of 140 trials (14%) were removed, with 868 trials remaining for the analysis.

The preliminary HLM models indicated that random subject and item effects improved the fit of the model, therefore, main effects (Counterbalancing Group, Congruency, Prime Compatibility, Word Ambiguity, and Word Frequency) were added as the predictor variables. After confirming that the addition of these variables improved the fit of the model and were accounting for RT variance not explained by the random subject and item effects, all two- and three-way interactions were added. However, even

[42] The final models including all variables (significant and non-significant) are presented in Appendix K.
[43] Of these 97 trials, 24 were false starts of the correct response and 73 were incorrect responses.

after a step-wise reduction of the non-significant interaction variables, the interaction model did not prove a better fit to the data than the main effects model. Therefore, the final model included only the main effects. Word Frequency and Congruency were significant and Word Ambiguity approached significance, as seen in Table 21, and Prime Compatibility and Counterbalancing Group were not significant. The Word Frequency and Congruency main effects will be addressed first, followed by a discussion of the similarities and differences in results to those of Alario et al. (2004). The main effect of Word Ambiguity will be discussed at the end of the section.

Table 21

FNS Final Model: Significant Main Effects

Variable	df	F-value	Sig.
Word Frequency	(1, 44)	9.865	$p < .01$
Congruency	(1, 801)	6.282	$p < .01$
Word Ambiguity	(2, 46)	3.173	$p = .051$

The significant main effect of Word Frequency indicates that, as expected, participants were faster to name pictures of high frequency nouns than low frequency nouns, regardless of the congruency condition. Participants also demonstrated sensitivity to the prime-target gender congruency, with fastest RTs in the neutral condition (M = 857ms, SD = 97ms), followed by the congruent condition (M = 869ms, SD = 105), and with slowest RTs in the incongruent condition (M = 900ms, SD =104ms). A post-hoc one-tailed paired samples t-test showed that RTs in the incongruent condition were significantly slower than RTs in the neutral condition, t (19) = 1.928, $p < .05$, and the difference between RTs in the congruent and incongruent conditions approached significance, *t* (19) = -1.654, *p* = .057. The difference between the congruent and neutral conditions was not significant. Mean RTs for the congruency conditions are presented in Figure 6.

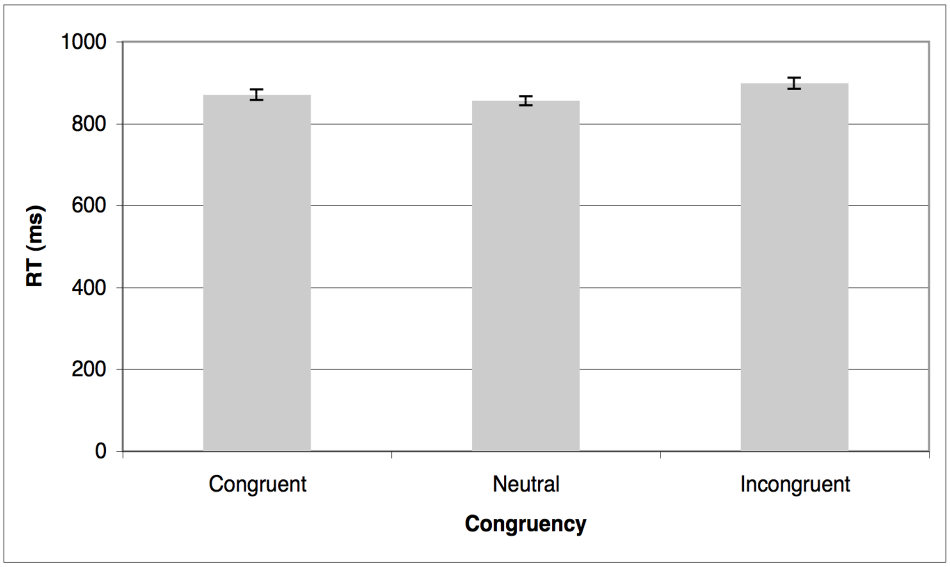


Figure 6. FNS RTs (ms) in congruent, neutral, and incongruent conditions

Despite a significant main effect of Congruency, the results indicate that only interference occurred, as RTs in the incongruent condition were slower than RTs in neutral and congruent conditions, but no difference was found between RTs in the neutral and congruent conditions. These findings do not replicate those of Alario et al. (2004), who found RTs in the congruent condition to be faster than RTs in the neutral condition (facilitation), and RTs in the incongruent condition to be slower than RTs in the neutral condition (interference). There are several possible explanations for this difference. First, unlike the primes used in Alario et al.'s study, the current study included syntactically incompatible prime-target combinations (*il/elle* prime + picture). As discussed in Chapter 3, it may be that Alario et al.'s results are due to determiner-noun pair co-occurrence frequency effects rather than gender priming effects, and that the syntactically incompatible prime-target combinations in the current task are masking these effects. To address this question, a post-hoc repeated-measures ANOVA was conducted with only syntactically compatible prime-target trials. Although the ANOVA revealed a significant main effect of congruency, $F\ (2, 40) = 5.114$, $p < .05$, a set of post-hoc paired-samples t-tests to compare RTs between the three congruency conditions still showed significant differences between congruent and incongruent RTs, $t\ (20) = -2.546$, $p < .05$, and neutral and incongruent RTs, $t\ (20) = -2.848$, $p = .01$ but not between congruent and neutral RTs, $t\ (20) = .690$, $p = .498$. That is, RTs in the congruent and neutral conditions remained similar, and both were faster than the RTs in the incongruent condition, even without the presence of the syntactically incompatible primes.

Furthermore, evidence that the syntactically incompatible primes are not simply masking a determiner-noun pair co-occurrence frequency effect comes from comparing RTs on the trials with syntactically incompatible vs. compatible primes. If co-occurrence frequency effects were playing a role, we would expect to see slower RTs for incompatible prime-target combinations in both the congruent and incongruent conditions because subject pronouns and nouns never occur together. However, in a set of paired-samples t-tests, there were no significant differences between syntactically compatible and syntactically incompatible primes in either the congruent, $t\ (20) = -.153$, $p = .880$, or incongruent, $t\ (20) = 1.389$, $p = .180$, conditions.

A second possible explanation for the similar congruent and neutral condition RTs is the inclusion of only one neutral prime (*chaque*) as opposed to eight possible primes (*le, la, un, une, mon, ma*, *il*, *elle*) in the congruent condition. In other words, participants saw the prime *chaque* 16 times during the task, whereas the other primes each occurred only twice. It may be that a practice effect within the task enabled the participants to respond quickly to the neutral trials with *chaque*, masking a facilitation effect of the congruent primes over the neutral prime. However, if this is the case, it is unclear why Alario et al. (2004) did not find the same pattern, given that they had a similar distribution of primes, with *chaque* as the only prime in the neutral condition.

A final possible explanation lies in the target pictures. Due to the restrictions of controlling for noun-ending ambiguity and avoiding L1-L2 cognate effects, only 16 of the target pictures used in Alario et al. (2004) were used in the current task, and 32 new pictures were added. The new pictures were taken from the same source (Snodgrass & Vanderwart, 1980), had high name agreement (between 93-100%), and had a higher mean frequency (51.7 per million) than the mean frequency of the target pictures used in Alario et al. (30 per million). To determine whether the picture selection is driving the differences in results in the current task and in that of Alario et al., a repeated-measures ANOVA for RT with Congruency as the within-subject variable was conducted including only trials with primes and target pictures that were used in both the current task and in Alario et al. If findings similar to Alario et al. were found when only looking at the same materials as used in their task, it would indicate that the different materials were responsible for the different result; however, no significant differences between the congruency conditions were found in this analysis, F (2, 40) = 1.286, $p > .288$.

Furthermore, the pattern of RTs was consistent with the findings in the current study, with fastest RTs in the neutral condition and slowest RTs in the incongruent condition.

Overall, although the results in the current task were not identical to those of Alario et al. (2004), the following conclusions can be made. First, a gender prime sensitivity was found, as demonstrated by faster RTs in the congruent condition than the incongruent condition. This result does not seem to be due to frequency co-occurrence effects because (a) RTs in the neutral condition were not slower than in the congruent condition despite the lower frequency of the neutral prime (although it is unclear whether the inclusion of a sole neutral prime is responsible for this result), and (b) the syntactically incompatible prime-target combinations, which never occur, were not slower than their syntactically compatible counterparts.

Second, the main difference between the current task design and that of Alario et al. (2004), the inclusion of syntactically incompatible primes, is not driving the difference in results. The presence of the syntactically incompatible primes in the task does not prevent priming effects from occurring.

Third, the presence of the syntactically incompatible primes in the current task provides a more complete understanding of the role of frequency co-occurrence effects, which helps to resolve the question raised in Alario et al.'s (2004) study. Specifically, Alario et al.'s symmetrical results (congruent RTs < neutral RTs < incongruent RTs) could have been explained as prime-target combination frequency effects rather than gender priming effects; however, the results of the current task indicate that prime-target frequency effects are not driving the results. Rather, a sensitivity to gender primes is causing faster RTs in the congruent condition than in the incongruent condition.

To summarize the congruency results, the FNSs showed gender priming effects in that they were faster to name a target picture when the prime was the same grammatical gender as the picture than when the prime's gender was different. This result supports the theory that a gender prime activates the appropriate gender node, which in turn, facilitates production of nouns also linked to that gender node.

In addition to main effects of Word Frequency and Congruency, the main effect of Word Ambiguity approached significance. Regardless of the congruency condition, FNSs were slower to produce target nouns with phonologically ambiguous endings than target nouns with unambiguous endings, or nouns that are exceptions. This finding is especially interesting as Levelt's (1989) production model maintains that a lemma's phonological form is activated after all syntactic information is made available. In other words, a target noun's phonological form should not influence production RT in a gender priming task, yet the RTs in the current task may indicate that the FNSs were slower to produce nouns with ambiguous endings, as shown in Figure 7.

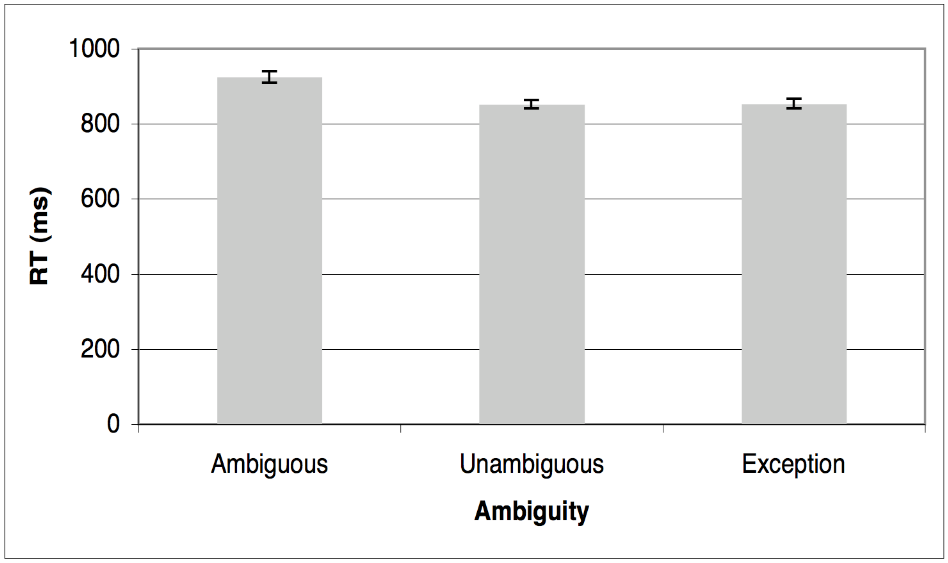


Figure 7. FNS RTs (ms) for ambigous, unambiguous, and exception target nouns

Before concluding that the slower RTs for ambiguous nouns represent a reliance on phonology during production, other potential explanations must be considered. First, characteristics of the target nouns that could affect picture naming RTs were compared. Table 22 presents the means for word frequency, name agreement, image agreement, image familiarity, and image complexity for ambiguous, unambiguous, and exception target nouns; in addition, the number of nouns with an unvoiced onset and the number of vowel-initial nouns are presented for each noun type.

Table 22

Characteristics of Ambiguous, Unambiguous, and Exception Target Nouns

	Ambiguous	Unambiguous	Exception
Word frequency (per million)	47.6 (2.8-504.3)	71.5 (1.4-479.9)	44.9 (2.3-306.5)
Name agreement	98.9 (93-100)	99.5 (93-100)	99.3 (96-100)
Image agreement	3.6 (2.4-4.7)	3.5 (2.5-4.6)	3.7 (2.9-4.4)
Image familiarity	3.3 (1.5-4.9)	3.7 (2.0-5.0)	3.6 (1.8-4.9)
Image complexity	2.8 (1.2-4.6)	2.6 (1.0-4.9)	2.9 (1.6-4.4)
Number (%) of words with unvoiced onset	7 of 16 (44%)	9 of 20 (45%)	10-12 (83%)
Number (%) of vowel-initial words	7 of 16 (44%)	0 of 20 (0%)	0 of 12 (0%)

Based on this comparison, the only potential factor that could explain the slower RTs for target nouns with ambiguous endings is the presence of vowel-initial words, of which there are seven in the ambiguous condition, but none in either the unambiguous or exception conditions. It is conceivable that weaker gender node-lemma links exist for vowel-initial words as opposed to consonant-initial words due to the fact that vowel-initial words occur less frequently with gender-marked determiners. For example, *ampoule* (light bulb, fem.) occurs with *l'* and *mon* rather than with the feminine forms, *la* and *ma*. Therefore, if gender node-lemma links are strengthened through determiner-noun occurrences, then in a priming task, the determiner prime, regardless of the congruency condition, may affect noun production differently for vowel-initial nouns. However, when the RTs for the vowel-initial nouns were compared to the RTs for the consonant-initial nouns in the ambiguous condition, the vowel-initial nouns had a faster mean RT (920ms) than the consonant-initial nouns (964ms), indicating that the presence of vowel-initial nouns is not driving the overall slower RTs for ambiguous target nouns.

A second potential explanation for the slower RTs for ambiguous nouns is the distribution of primes across the ambiguous, unambiguous, and exception conditions. If certain primes result in faster picture naming RTs than others, and the distribution of primes is skewed such that the nouns in the ambiguous condition are more frequently primed by the "slower" primes, this could affect overall RTs for ambiguous nouns. Figure 8 shows that, overall, *mon/ma* primes result in slower RTs than the other primes.

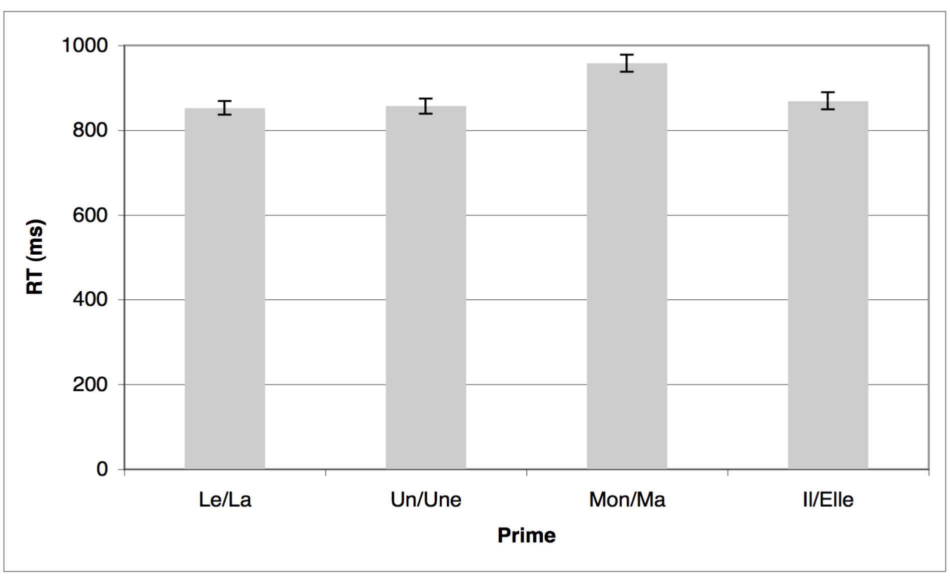


Figure 8. Mean RTs (ms) for *le/la*, *un/une*, *mon/ma*, and *il/elle* primes

Given that definite and indefinite determiners (*le/le* and *un/une*) are devoid of semantic meaning, whereas possessive pronouns (*mon/ma*) carry meaning, it may be that processing the possessive pronoun primes takes longer, and, therefore, RTs are slower. Furthermore, *mon* does not always indicate masculine gender as it is the form used before vowel-initial nouns, regardless of the noun's gender (e.g., *mon ami* [my friend, masc.], *mon amie* [my friend, fem.]). In other words, *mon*, like *l'*, serves as the neutralized form for vowel-initial nouns. Looking at the distribution of primes for the ambiguous, unambiguous, and exception nouns (Table 23), the nouns in the ambiguous condition were primed by *mon/ma* more often than nouns in the unambiguous and exception conditions. That is, the ambiguous nouns have a higher percentage of "slower" primes than the unambiguous and exception nouns.

Table 23

Distribution of Primes Across Ambiguous, Unambiguous, and Exception Target Nouns

	Ambiguous	Unambiguous	Exception
Le/La	4%	27%	17%
Un/Une	24%	7%	23%
Mon/Ma	22%	17%	12%
Il/Elle	17%	18%	15%

To explore the possibility of the *mon/ma* primes contributing to the ambiguity effect, a follow-up analysis was conducted in which all trials with *mon/ma* primes were removed. However, the RT pattern for ambiguous, unambiguous, and exception nouns remained the same, with slower RTs for the ambiguous nouns. The higher percentage of *mon*/*ma* primes for ambiguous nouns does not account for the slower RTs.

The remaining potential explanation for the slower RTs for ambiguous nouns is that FNSs rely on phonological noun endings, at least to some degree, during this task. Considering a noun's phonological ending serves as a gender predictor, it is logical to conclude that the main effect of Word Ambiguity is somehow related to gender activation. Therefore, either this particular task contains an unintended component in which the FNSs are activating and relying on the phonological noun ending during production, or, gender processing in French integrates phonological form, despite Levelt's (1989) claim that phonological form is independent of gender processing. Because previous gender priming research has not considered noun ending ambiguity, it is impossible at this point to know whether this effect generalizes to other L1 language groups and priming tasks, or whether it is an effect specific to French (or perhaps Romance languages), or only this particular task. At any rate, because the ambiguous, unambiguous, and exception nouns are counterbalanced across the congruency conditions, and Word Ambiguity does not interact with Congruency, the finding of slower RTs for ambiguous nouns does not interfere with the conclusion that the FNSs are showing gender priming effects. It will be interesting, however, to compare the role of ambiguity among the NNS language groups to determine whether NNSs perform similarly to FNSs in picture naming.

8.4.2 Spanish Native Speakers

Of 1776 trials, 648 (36%) were removed due to false starts and inaccurate responses,[44] 81 (5%) were removed due to incorrect gender assignment on the gender assignment post-test, 85 (5%) were trimmed due to RTs faster than 400ms or slower than 2000ms, and finally, 6 (.3%) were removed as outliers. Overall, a total of 820 (46%) of 1776 trials were removed, with 956 trials remaining for analysis.

The first HLM analysis carried out for the SNSs was the same as that for the FNSs. The preliminary models indicated that random subject and item effects improved the fit of the model, therefore, main effects (Counterbalancing Group, Congruency, Prime Compatibility, Word Ambiguity, and Word Frequency) were added as predictor variables. After confirming that the addition of these variables improved the fit of the model and were accounting for RT variance not explained by the random subject and item effects, all two- and three-way interactions were added. A step-wise reduction of the non-significant interaction variables resulted in the final model, presented in Table 24, which proved a better fit to the data than the main effects model.

[44] Of these 648 trials, 48 were false starts of the correct response and 600 were incorrect responses.

Table 24

SNS First Analysis Final Model: Significant Main Effects and Interactions

Variable	df	F-value	Sig.
Word Frequency	(1, 410)	5.799	$p < .05$
Congruency+	(1, 868)	6.531	$p < .05$
Counterbalancing Group x Congruency	(4, 84)	3.295	$p < .05$
Prime Compatibility x Word Ambiguity+	(4, 580)	3.763	$p < .01$
Prime Compatibility x Word Ambiguity x Word Frequency*	(4, 586)	3.633	$p < .01$

+Subsumed by higher order interactions
*Interactions discussed in Appendix K

As with the FNSs, main effects of Word Frequency and Congruency were found. However, both of these main effects, as well as a two-way interaction between Prime Compatibility and Word Ambiguity, were subsumed by higher order interactions, and, thus, were not considered independently. Because the primary goal of this analysis was to determine whether participants show gender priming effects and the variables that potentially predict these effects, interactions between variables that do not include Congruency are be not be discussed in the main text, but rather addressed in Appendix K. Furthermore, the higher order interactions often have either too few people or too few items to be meaningful, and it is important not to overinterpret the results.

The Counterbalancing Group by Congruency interaction indicates that participants showed differing priming effects based on the group to which they were randomly assigned. Because Group A always saw List 1 in the congruent condition, Group B always saw List 1 in the neutral condition, etc., the Counterbalancing Group by Congruency interaction may be representative of a list effect, as shown in Figure 9.

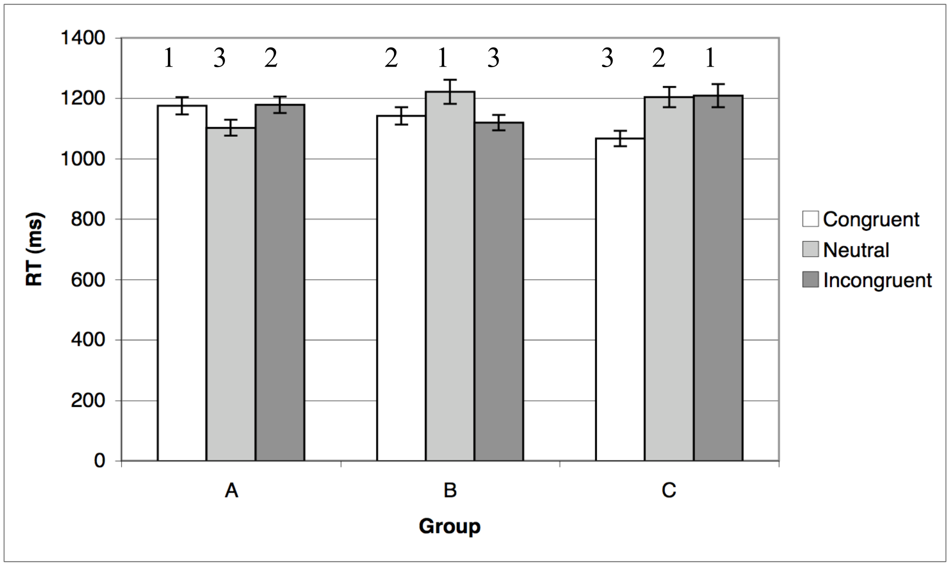


Figure 9. SNS Counterbalancing Group by Congruency interaction, with list (1, 2, 3) displayed above each column to indicate which Group saw which list in each Congruency condition

An in-depth examination of possible list and/or group effects is addressed in Appendix K, but on the whole, it is impossible to determine how much of the interaction is due to list effects and how much is due to differences across the three groups. At most the data showing faster RTs in the congruent condition than in the incongruent condition for Group C provide only very weak evidence for priming.

Based on this first analysis, it is unclear how robust the gender priming effects are for the SNSs. To explore the role of AO, the number of years a participant has known French, and the number of years spent in France, a second analysis was carried out including these variables as predictors. The goal of this second analysis was to investigate the role of these language experience factors in gender priming. In other words, the second analysis addresses the question: does gender priming occur for some of the participants depending on their French language learning experience?

The preliminary HLM models in the second analysis indicated that random subject and item effects improved the fit of the model, as did adding in the main effects, and finally the interactions. The final model after removing non-significant interactions is presented in Table 25.

Table 25

SNS Second Analysis Final Model: Significant Main Effects and Interactions

Variable	df	F-value	Sig.
Word Frequency	(1, 37)	15.116	$p < .01$
Counterbalancing Group+	(2, 34)	4.042	$p < .05$
AO+	(1, 35)	20.923	$p < .01$
Years Known+	(1, 34)	14.546	$p < .01$
Counterbalancing Group x AO+	(2, 35)	4.128	$p < .05$
Counterbalancing Group x Years Known+	(2, 34)	4.805	$p < .05$
Counterbalancing Group x Word Ambiguity*	(4, 866)	2.545	$p < .05$
Prime Compatibility x Word Ambiguity+	(2, 808)	4.449	$p < .05$
Prime Compatibility x Years France+	(2, 869)	3.179	$p < .05$
AO x Years Known+	(1, 34)	15.847	$p < .01$
Word Ambiguity x Years France+	(2, 861)	4.484	$p < .05$
Congruency x Word Ambiguity x Years Known	(1, 859)	3.02	$p < .05$
Prime Compatibility x Word Ambiguity x Years France*	(4, 862)	2.753	$p < .05$
Prime Compatibility x Years France x Years Known*	(2, 866)	3.941	$p < .05$
Counterbalancing Group x Years France x Years Known*	(2, 35)	4.122	$p < .05$
Counterbalancing Group x AO x Years Known*	(2, 35)	3.963	$p < .05$

+Subsumed by higher order interactions
*Interactions discussed in Appendix K

A main effect of Word Frequency indicates that, similar to FNSs, SNS participants were faster to name pictures of high frequency nouns than low frequency nouns. In addition, the main effect of Congruency found in the first analysis disappeared, but a three-way interaction between Congruency, Word Ambiguity, and Years Known was significant. In order to visually represent this interaction (Figures 10-12), the continuous variable, Years Known, was divided into three categorical groups, 2-9 years, 10-14 years, and 15-23 years; the divisions were selected on the basis of creating equal groups, to the extent possible. However, before interpreting this interaction, it is important to note the small subject size within each Years Known group, with only 13 participants in the 2-9 year group, 13 in the 10-14 year group, and 11 in the 15-23 year group.

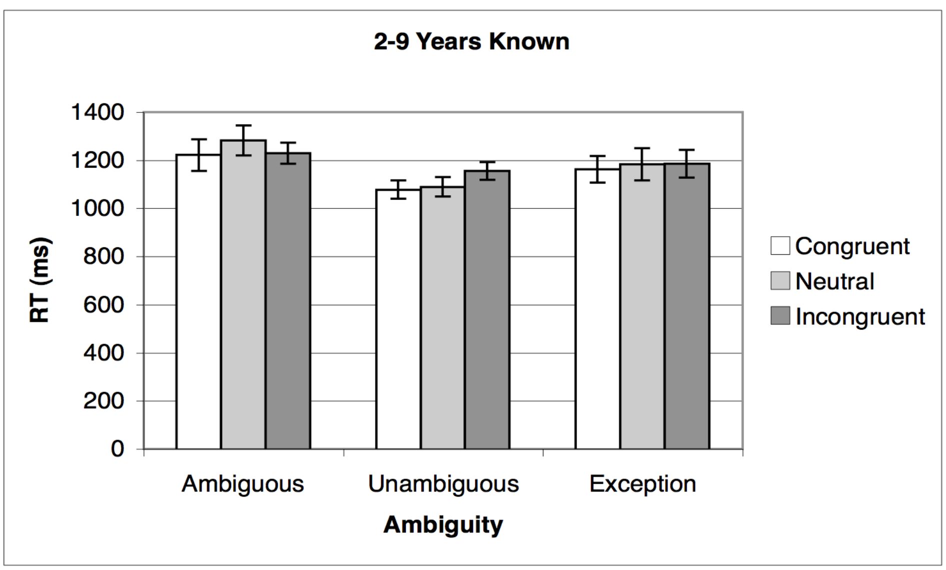


Figure 10. SNS interaction between Congruency and Word Ambiguity for participants who have known French for 2-9 years

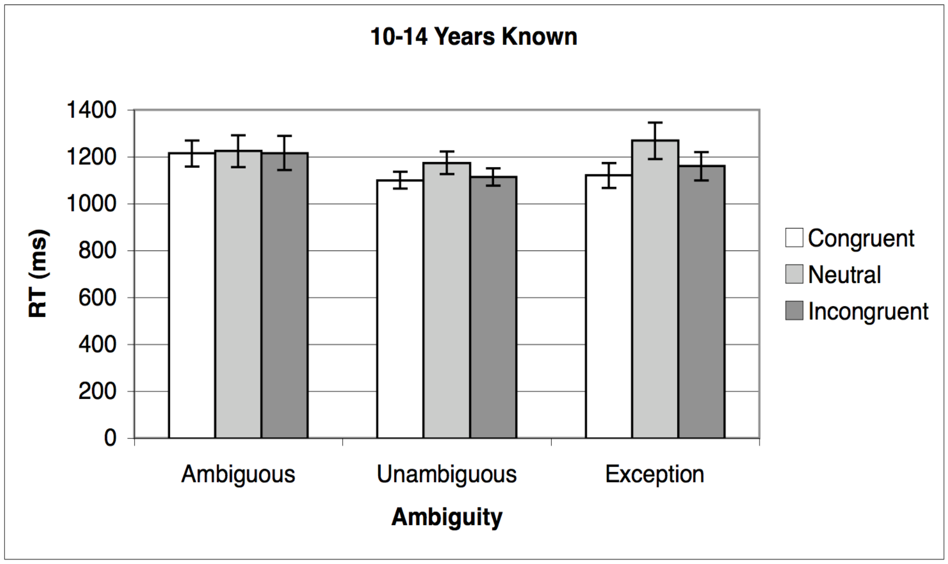


Figure 11. SNS interaction between Congruency and Word Ambiguity for participants who have known French for 10-14 years

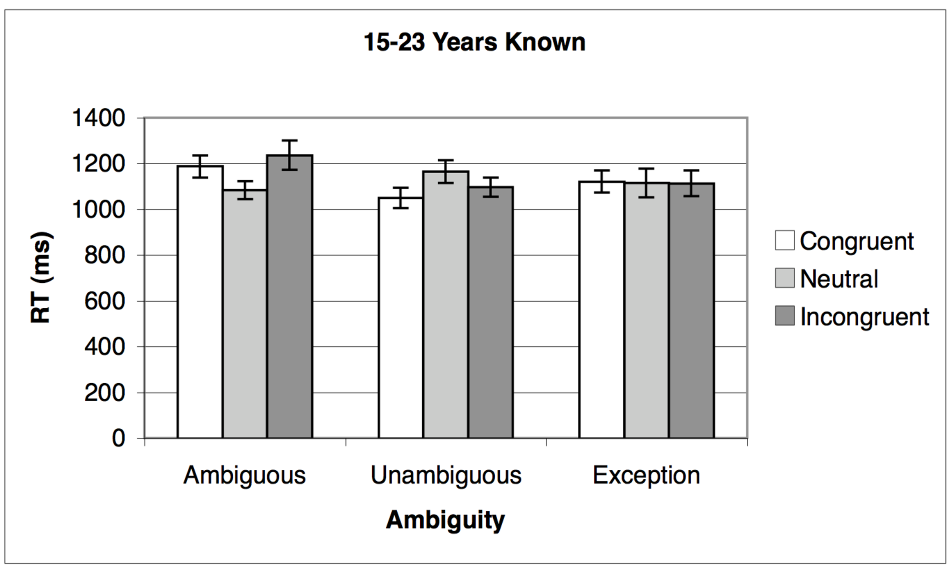


Figure 12. SNS interaction between Congruency and Word Ambiguity for participants who have known French for 15-23 years

For the participants who have known French the least amount of time (2-9 years), there is little difference in congruency RTs for the ambiguous and exception nouns, but slightly slower RTs in the incongruent condition for unambiguous nouns, suggesting a priming effect for unambiguous nouns. However, with only 13 participants in this group, it is difficult to draw a definitive conclusion regarding priming effects. For the participants who have known French for 10-14 years and 15-23 years, there is no evidence for priming effects based on Word Ambiguity. That is, other than a slight RT disadvantage for unambiguous nouns in the incongruent condition for the participants who have known French for 2-9 years, a clear gender priming pattern based on Word Ambiguity does not emerge for any of the years known groups. The remaining higher order interactions that do not involve Congruency are discussed in Appendix K.

To conclude, SNSs show, at most, only weak gender priming effects, which are likely to be driven by list effects. Even when the potential language experience factors, AO, number of years spent in France, and number of years a participant has known the language, are added into the model as predictors, consistent priming effects do not emerge.

8.4.3 Dutch Native Speakers

Of 1824 trials, 654 (36%) were removed due to false starts and incorrect responses,[45] 107 (6%) were removed due to incorrect gender assignment on the gender assignment post-test, 96 (5%) were removed due to RTs that were either faster than 400ms or slower than 2000ms, and 2 trials (.1%) were removed as outliers. Overall, a total of 859 (47%) of the 1824 trials were removed, with 965 trials remaining for analysis.

---

[45] Of these 654 trials, 36 were false starts of the correct response and 618 were incorrect responses.

The first HLM analysis carried out for the DNSs was the same as that for the FNSs. The preliminary models indicated that random subject and item effects improved the fit of the model, therefore, main effects (Counterbalancing Group, Congruency, Prime Compatibility, Word Ambiguity, and Word Frequency) were added as the predictor variables. After confirming that the addition of these variables improved the fit of the model and were accounting for RT variance not explained by the random subject and item effects, all two- and three-way interactions were added. A step-wise reduction of the non-significant interaction variables resulted in the final model, presented in Table 26, which proved a better fit to the data than the main effects model.

Table 26

DNS First Analysis Final Model: Significant Main Effects and Interactions

Variable	df	F-value	Sig.
Word Frequency+	(1, 191)	6.288	$p < .05$
Counterbalancing Group x Word Frequency+	(2, 754)	5.656	$p < .01$
Prime Compatibility x Word Ambiguity+	(2, 264)	3.378	$p < .05$
Counterbalancing Group x Congruency x Word Ambiguity	(8, 107)	2.491	$p < .05$
Counterbalancing Group x Word Ambiguity x Word Frequency*	(4, 795)	3.753	$p < .01$
Prime Compatibility x Word Ambiguity x Word Frequency*	(4, 389)	3.463	$p < .01$

+Subsumed by higher order interactions
*Interactions discussed in Appendix K

A main effect of Word Frequency and two two-way interactions (Counterbalancing Group and Word Frequency, and Prime Compatibility and Word Ambiguity) were significant; however, as all were subsumed by higher order interactions, they were not considered independently.

There was no main effect of Congruency, indicating that the DNSs did not show priming effects. A three-way interaction between Counterbalancing Group, Congruency, and Word Ambiguity was significant, and is displayed in Figures 13-15. Further examination of the possible list and group effects is presented in Appendix K.

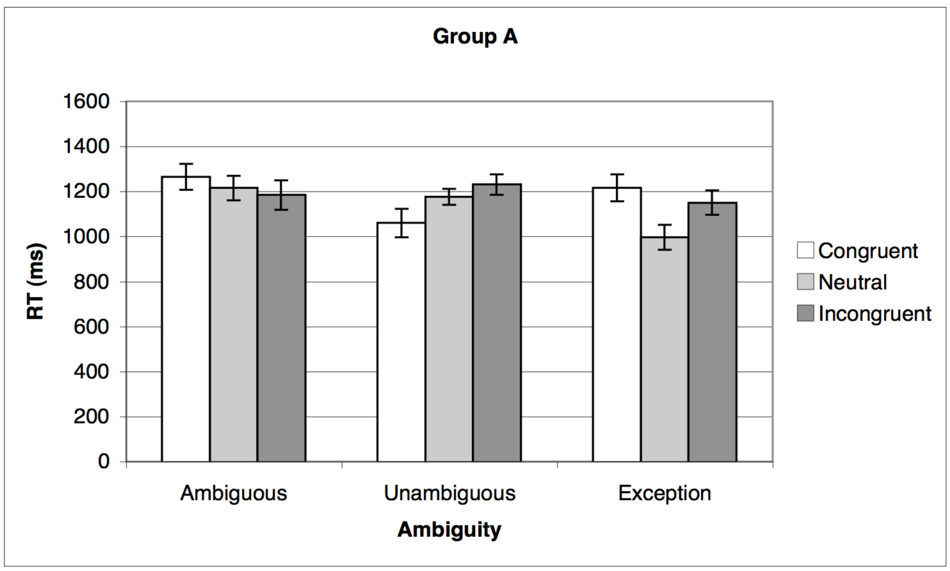


Figure 13. DNS interaction between Congruency and Word Ambiguity: Group A

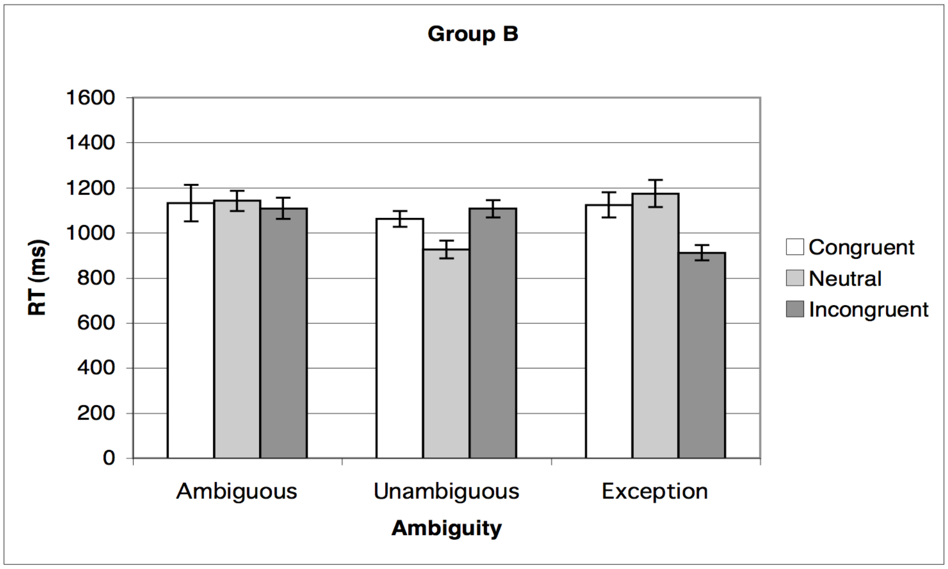


Figure 14. DNS interaction between Congruency and Word Ambiguity: Group B

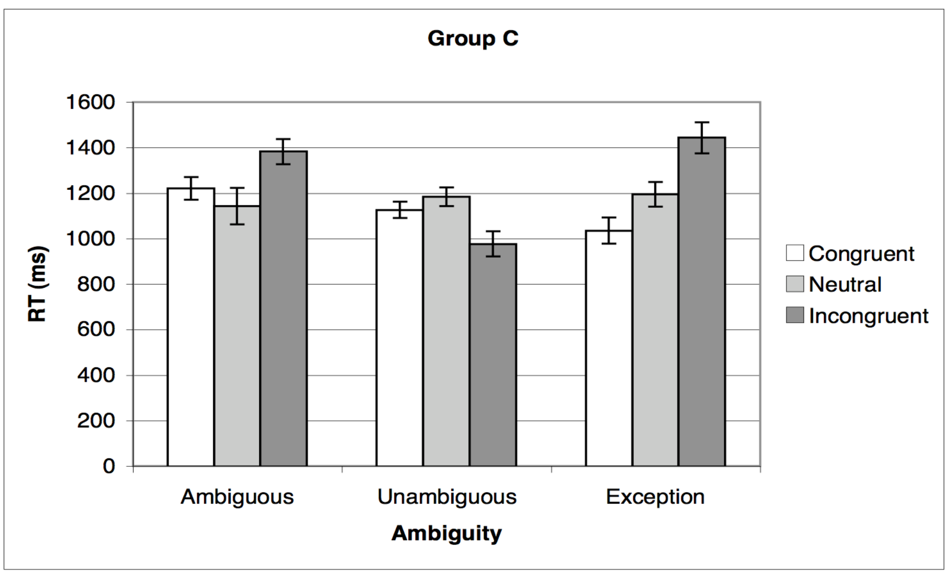


Figure 15. DNS interaction between Congruency and Word Ambiguity: Group C

For Group A participants, there was little difference in RT for ambiguous nouns, regardless of the congruency condition. For the unambiguous nouns, however, RTs were faster in the congruent condition than in the neutral and incongruent conditions, suggesting a priming effect. The RTs for exception nouns were faster in the neutral condition, although the similar RTs for exception nouns in the congruent and incongruent conditions indicate no priming effects are occurring. Group B participants showed faster RTs in the neutral condition for unambiguous nouns and faster RTs in the incongruent condition for exception nouns, but again, neither of these effects demonstrates gender priming effects as RTs were not faster in the congruent condition than the incongruent condition. Finally, Group C participants showed priming effects for both ambiguous and exception nouns, with faster RTs in the congruent condition than in the incongruent condition. Conversely, for unambiguous nouns, RTs were fastest in the incongruent condition.

Whereas priming effects are apparent for Group A (for unambiguous nouns) and Group C (for ambiguous and exception nouns), based on the small number of participants per group (14 participants in Group A, and 12 participants each in Groups B and C), the lack of priming effects for Group B, and the differing priming effects for Group A and Group C, it is not possible to establish a pattern of priming effects based on Word Ambiguity. However, that the DNS participants show a sensitivity to ambiguity is noteworthy, and will be addressed in more detail in Section 8.4.5.

The remaining significant interactions (Counterbalancing Group, Word Ambiguity and Word Frequency, and Prime Compatibility, Word Ambiguity, and Word Frequency) do not contribute to the discussion of gender priming effects, and will, therefore, be addressed in Appendix K.

This first analysis demonstrates that, overall, DNSs do not show gender priming effects. The second analysis considered AO, the number of years spent in France and the number of years a participant has known French as predictor variables to determine whether gender priming effects occur based on these additional factors. The preliminary HLM models in the second analysis indicated that random subject and item effects improved the fit of the model, as did adding in the main effects, and finally the interactions. The final model after removing non-significant interactions is presented in Table 27.

Table 27

DNS Second Analysis Final Model: Significant Main Effects and Interactions

Variable	df	F-value	Sig.
Word Frequency+	(1, 69)	41.834	$p < .01$
Congruency+	(1, 891)	9.0	$p < .01$
Word Ambiguity+	(2, 893)	3.945	$p < .05$
Counterbalancing Group x Congruency+	(4, 339)	4.977	$p < .01$
Congruency x AO	(2, 893)	5.341	$p < .01$
Congruency x Years Known+	(2, 888)	3.345	$p < .05$
Congruency x Years France+	(2, 887)	6.383	$p < .01$
Word Ambiguity x Years Known+	(2, 890)	5.418	$p < .01$
Word Ambiguity x AO+	(2, 889)	4.317	$p < .05$
Counterbalancing Group x Congruency x Years Known	(4, 891)	4.331	$p < .01$
Counterbalancing Group x Word Ambiguity x Years France*	(4, 888)	3.652	$p < .01$
Counterbalancing Group x Word Ambiguity x Years Known*	(4, 889)	2.577	$p < .05$
Congruency x Years France x Years Known	(2, 885)	4.962	$p < .01$
Word Frequency x AO x Years Known*	(1, 885)	11.6	$p < .01$
Word Ambiguity x AO x Years Known*	(2, 890)	5.459	$p < .01$

+Subsumed by higher order interactions
*Interactions discussed in Appendix K

Main effects of Word Frequency, Congruency, and Word Ambiguity, and two-way interactions between Counterbalancing Group and Congruency, Congruency and Years Known, Congruency and Years France, Word Ambiguity and Years Known, and Word Ambiguity and AO, were all significant. However, as each of these was subsumed by a higher order three-way interaction, they were not considered independently. Furthermore, the interactions that do not include Congruency do not contribute to the discussion of priming effects, and will be addressed in Appendix K.

A significant two-way interaction between Congruency and AO is shown in Figure 16, with the continuous variable, AO, transformed into three categorical groups, 8-10 years (21 participants), 11 years (8 participants), and 12-13 years (9 participants); the divisions were selected on the basis of creating equal groups, to the extent possible.

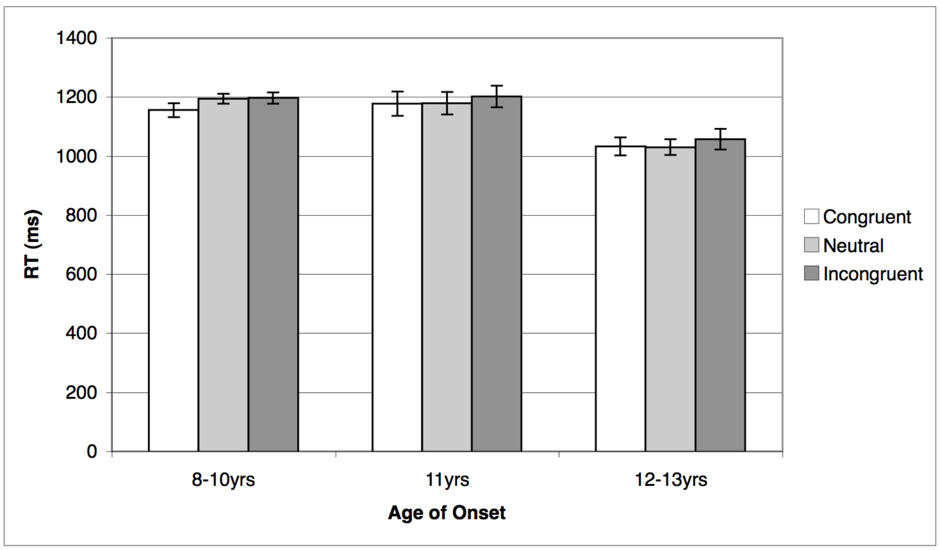


Figure 16. DNS interaction between Congruency and AO

Although the difference in RTs between congruent, neutral, and incongruent conditions was minimal for each group, the participants with the earliest AO showed faster RTs in the congruent condition as compared to the neutral and incongruent conditions, suggesting that priming effects may occur for learners who began learning French between 8-10 years of age. However, the overall early and uniform AO for the DNSs makes it difficult to examine this relationship, especially considering the uneven number of participants in the three AO groups. Although it appears that priming effects may occur for learners with an early AO, that the 11 and 12-13 year AO groups show no priming effects indicate that, despite their overall early AO, priming effects are not occurring for the DNSs.

A significant three-way interaction between Counterbalancing Group, Congruency, and Years Known was found. In order to interpret this interaction, the continuous variable, Years Known, was transformed into a categorical variable with three groups: 10-11 years, 12-17 years, and 18-50 years. The divisions were selected on the basis of creating equal groups, to the extent possible. However, due to the small number of participants in each cell (range 3-6), this interaction is difficult to interpret and would have limited bearing on the discussion on gender priming. The same is true for the three-way interaction between Congruency, Years France and Years Known. Therefore, these variables will be considered later in a correlation matrix in an attempt to determine whether they account for the presence or absence of gender priming effects. The remaining significant interactions do not contribute to the discussion of gender priming effects, and will, therefore, be addressed in Appendix K.

The second HLM analysis for DNSs shows that gender priming effects are not occurring. Of the additional language experience factors, only AO emerged as a possible variable contributing to NNS priming effects; however, based on the overall early and uniform AO for the DNS participants, it is difficult to draw a definitive conclusion about the role of AO in gender priming. This relationship will be further explored in the correlation matrix.

8.4.4 English Native Speakers

Of 2016 trials, 472 (23%) were removed due to false starts and incorrect responses,[46] 167 (8%) were removed due to incorrect gender assignment on the gender assignment post-test, 63 (3%) were removed due to RTs faster than 400ms or slower than 2000ms, and 10 (.5%) were removed as outliers. Overall, a total of 712 (35%) of 2016 trials were removed, with 1304 trials remaining for analysis.

The preliminary models of the first HLM analysis indicated that random subject and items effects improved the fit of the model, therefore, main effects (Counterbalancing Group, Congruency, Prime Compatibility, Word Ambiguity, and Word Frequency) were added as predictor variables. These variables improved the fit of the model, and all two- and three-way interactions were added. A step-wise reduction of the non-significant interaction variables resulted in the final model that was a better fit to the data than the main effects model. The final model is presented in Table 28.

Table 28

ENS First Analysis Final Model: Significant Main Effects and Interactions

Variable	df	F-value	Sig.
Word Frequency	(1, 42)	15.281	$p < .01$
Prime Compatibility	(1, 1080)	3.997	$p < .05$
Counterbalancing Group x Congruency	(4, 287)	3.164	$p < .05$

Main effects of Word Frequency and Prime Compatibility, and a two-way interaction between Counterbalancing Group and Congruency were found. Because there was no

[46] Of these 472 trials, 90 were false starts of the correct response and 382 were incorrect responses.

main effect of Congruency, the two-way interaction between Counterbalancing Group and Congruency and a discussion of list and/or group effects is presented in Appendix K. The main effect of Word Frequency indicates that ENS participants were faster at naming pictures of high frequency words than low frequency words. The main effect of Prime Compatibility, as shown in Figure 17, demonstrates that the ENS were fastest at naming pictures with an incompatible prime (*M* = 1081ms, *SD* = 158ms), followed by compatible primes (*M* = 1105ms, *SD* = 107ms), and slowest at naming pictures with a neutral prime (*M* = 1122 ms, *SD* = 120ms), regardless of the congruency condition.

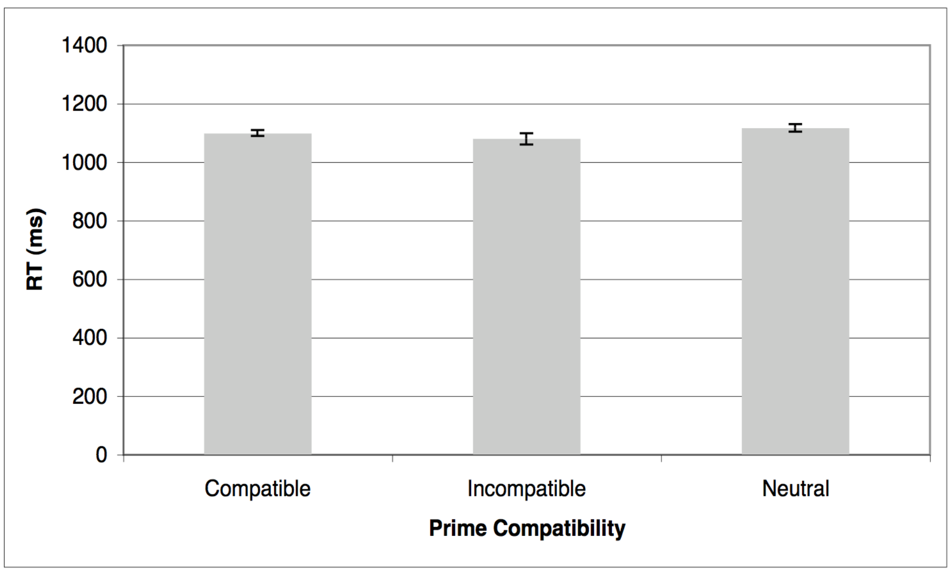


Figure 17. ENS RTs (ms) for compatible, incompatible, and neutral primes

To explore the role of AO, number of years spent in France, and number of years participant has known French, the second analysis was carried out with these factors added as predictor variables. The final model is presented in Table 29.

Table 29

ENS Second Analysis Final Model: Significant Main Effects and Interactions

Variable	df	F-value	Sig.
Counterbalancing Group+	(2, 46)	14.666	$p < .01$
Word Frequency+	(1, 60)	21.218	$p < .01$
Years in France+	(1, 42)	13.143	$p < .01$
Years Known+	(1, 42)	13.261	$p < .01$
Counterbalancing Group x Congruency+	(4, 1254)	3.131	$p < .05$
Counterbalancing Group x AO+	(2, 45)	15.374	$p < .01$
Counterbalancing Group x Years France+	(2, 41)	7.004	$p < .01$
Counterbalancing Group x Years Known+	(2, 42)	6.904	$p < .01$
Prime Compatibility x Years France	(1, 1226)	8.254	$p < .01$
AO x Years France+	(1, 41)	13.763	$p < .01$
AO x Years Known+	(1, 41)	16.679	$p < .01$
Years in France x Years Known+	(1, 42)	7.883	$p < .01$
Counterbalancing Group x Congruency x Years Known	(4, 1225)	3.927	$p < .01$
Counterbalancing Group x Congruency x Years France	(4, 1218)	4.207	$p < .01$
Counterbalancing Group x Congruency x AO	(6, 1220)	2.28	$p < .05$
Counterbalancing Group x AO x Years Known*	(2, 41)	7.638	$p < .01$
Counterbalancing Group x AO x Years France*	(2, 41)	5.991	$p < .01$
Counterbalancing Group x Years in France x Years Known*	(2, 41)	6.411	$p < .01$
Word Frequency x AO x Years Known*	(1, 1219)	11.091	$p < .01$
Word Frequency x Years France x Years Known*	(1, 1218)	5.749	$p < .05$

+Subsumed by higher order interactions
*Interactions discussed in Appendix K

Several main effects, and two- and three-way interactions were significant in the second HLM analysis; however, only the highest-order interactions are addressed.

Whereas Prime Compatibility emerged as a significant main effect in the first analysis, it is only significant in a two-way interaction with Years France in the second analysis. To examine this interaction, the continuous variable, Years France, was transformed into three categorical groups: 0-3 years (14 participants), 5-20 years (12 participants), and 21-41 years (16 participants); the divisions were selected on the basis of creating equal groups, to the extent possible. As shown in Figure 18, participants who have spent 5-20 years in France and 21-41 years in France showed similar RTs for compatible, incompatible, and neutral primes, but the participants who have spent 0-3 years in France showed faster RTs for the incompatible primes. In this case, adding Years France into the model accounted for the main effect of Prime Compatibility seen in the first model. That is, the effect of prime type depends on the number of years the participants had spent in France. Interestingly, the participants who had been in France the longest performed similarly to FNSs, who did not show any effect of Prime Compatibility, whereas the participants who had spent the least amount of time in France were less native-like in that they showed a difference in RT based on Prime Compatibility. This interaction does not, however, reveal priming effects based on either Prime Compatibility or Years France.

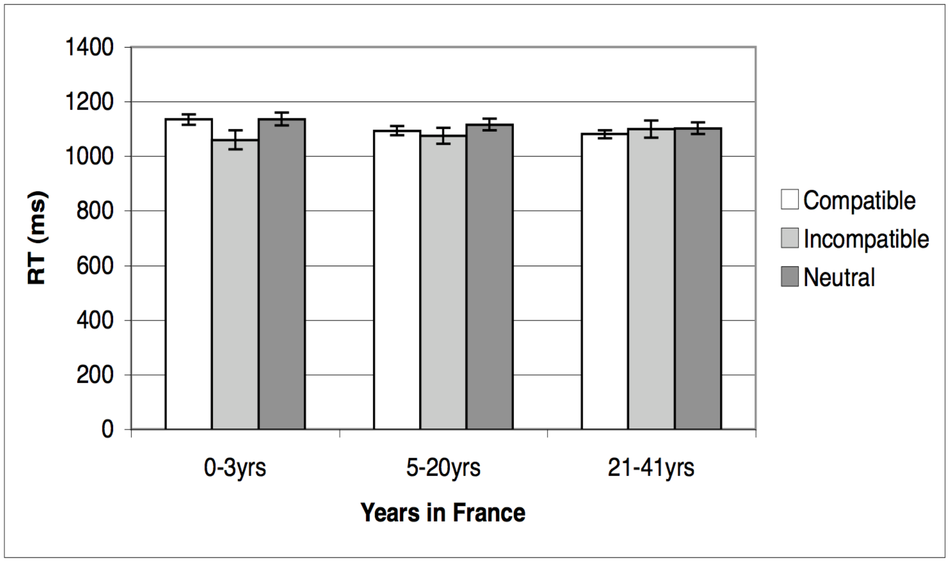


Figure 18. ENS interaction between Prime Compatibility and Years France

As discussed in Appendix K, list effects were most likely driving the Counterbalancing Group by Congruency interaction in the first analysis. The following significant interactions in the second analysis involve Counterbalancing Group and Congruency, as well as a third variable, and were also likely to be driven by list effects.

- Counterbalancing Group, Congruency, Years France
- Counterbalancing Group, Congruency, Years Known
- Counterbalancing Group, Congruency, AO

In other words, these interactions do not represent gender priming effects based on the number of years spent in France, the number of years a participant has known French, or AO; instead, they represent an interaction between these variables and list. For this reason, they are not addressed as meaningful outcomes in the gender priming analysis. The remaining significant interactions do not involve Congruency, and will, therefore, be addressed in Appendix K, as they do not bear on the primary focus of this analysis: gender priming effects.

8.4.5 Discussion

The goal of the gender priming analyses presented above was to determine whether any of the four language groups show gender priming effects. The task was designed to replicate Alario et al.'s (2004) study in which FNSs were presented with gender-marked or neutral determiner primes, followed by target pictures to be named. The prime-target pairs were congruent, incongruent or neutral. Their results showed clear priming effects, with faster RTs in the congruent condition as compared to the neutral condition, and slower RTs in the incongruent condition as compared to the neutral condition. However, it is possible that these results were a function of determiner-noun co-occurrence effects rather than priming effects. In an attempt to address this limitation, incompatible primes (*il/elle*) were included in the current task, as they never co-occur with nouns. In addition, noun ending ambiguity as an indicator of gender was included as a variable.

HLM was used for the analysis, with RT modeled as a function of Counterbalancing Group, Congruency, Prime Compatibility, Word Ambiguity, and Word Frequency as within-subjects fixed factors. The final model for the FNSs resulted in two significant main effects and no interactions. First, FNSs showed a Word Frequency effect they were faster at naming pictures of high frequency nouns than low frequency nouns. Second, the FNSs showed an effect of Congruency. Although the FNSs in the current study did not replicate the Alario et al. (2004) priming results (congruent RT < neutral RT < incongruent RT), the difference between RTs in the congruent and incongruent conditions nevertheless indicates gender priming is occurring. Finally, a main effect of

Word Ambiguity approached significance, with slower RTs for nouns with ambiguous endings. Although Word Ambiguity did not interact with Congruency, this result suggests that FNSs may rely on the phonological noun ending during picture naming, at least in this task.

The same HLM model was applied for the three NNS language groups. For the SNSs, priming effects were weak, at best, and confounded by Counterbalancing Group and/or list effects, making it difficult to definitively conclude that gender priming occurred. In an attempt to determine the role of AO, the number of years spent in France, or the number of years participants had known French, a second HLM analysis was carried out with these additional variables as predictors. The results from the second model showed a main effect of Word Frequency, as was found for FNSs, with faster RTs for high frequency nouns than low frequency nouns. In addition, a three-way interaction between Congruency, Word Ambiguity, and Years Known was found; however, a closer look at the data revealed no difference in RTs between the congruent and incongruent conditions based on either Word Ambiguity or Years Known. Therefore, the SNSs did not show any convincing evidence of gender priming, even when the additional language learning factors were included in the model.

The first HLM model for the DNSs revealed a significant three-way interaction for Counterbalancing Group, Congruency and Word Ambiguity. Although Group A showed gender priming effects for unambiguous nouns and Group C showed gender priming effects for ambiguous and exception nouns, the pattern was not consistent, and furthermore, Group B did not show any gender priming, regardless of noun ambiguity. Therefore, it is difficult to establish a general pattern of gender priming based on Word

Ambiguity. The second model including AO, Years Known, and Years France revealed a significant two-way interaction between Congruency and AO, with the earliest acquirers showing slightly faster RTs in the congruent condition. However, considering no main effect of Congruency was found in the first model, despite the DNSs having the earliest and most uniform AO of the three NNS groups, the two-way interaction in the second model is not convincing evidence of priming effects for the DNSs.

For the ENSs, the first HLM model revealed main effects of Word Frequency, with faster RTs for high frequency nouns than for low frequency nouns, and Prime Compatibility, with fastest RTs for incompatible primes and slowest RTs for neutral primes. However, a two-way interaction between Prime Compatibility and Years France was significant in the second model, indicating that only the participants who had been in France the least amount of time showed a difference between prime types. Similar to the SNS results, ENSs showed an interaction between Counterbalancing Group and Congruency, making it difficult to disentangle possible gender priming effects from group and list effects. At best, ENS priming effects were weak and inconsistent.

Considering only the first HLM model run for each group, which did not include NNS language experience factors, an interesting result for Word Frequency emerged. Both the FNS and ENS participants showed a main effect of Word Frequency, with faster RTs for higher frequency words, but the SNS and DNS participants did not. A potential explanation is that the ENSs have a larger vocabulary size than the SNS and DNS participants (based on the percent of trials removed in this task due to inaccurate responses), thereby increasing the frequency range of items included in the analysis and allowing for frequency effects to emerge. If, for the SNS and DNS participants, the target nouns produced in this task were limited to the higher frequency words, a frequency effect may not have appeared. To investigate this possibility, the distribution of low, medium, and high frequency words included in the analysis was compared across the NNS groups.[47] Because Word Frequency constitutes a continuous variable, it was divided into three categorical groups on the basis of creating equal groups, to the extent possible (low frequency = 1-12 per million, n = 17; medium frequency = 13-33 per million, n = 17; high frequency = 36-504 per million, n = 14). Percentages are presented in Table 30.

Table 30

Distribution (%) of Low, Medium, and High Frequency Words for Each NNS Language Group

	SNS	DNS	ENS
Low frequency	25%	25%	27%
Medium frequency	30%	32%	33%
High frequency	45%	43%	40%

Although the ENSs show a slightly more even distribution across frequency groups, the difference is not great enough to account for the lack of frequency effects for the SNS and DNS participants. In addition, the range of frequencies was the same for all three groups, as every word in the list was included for at least one person from each language group. Therefore, in terms of frequency effects in picture naming, it appears that the ENSs are performing in a more native-like fashion than the SNS and DNS participants. It

[47] It is important to keep in mind that frequency norms are based on NSs and it is difficult to determine whether these norms represent frequency for NNSs as well.

is possible that this effect is due to the significant difference in the mean number of years spent in France between the ENSs (15.6 years) and the SNSs and DNSs (3.1 years and 3.9 years, respectively).

In the second HLM model, however, which also included NNS language experience factors, the main effect of Word Frequency disappeared for the ENSs, but was significant for the SNSs. That is, once the language experience factors were added into the model and accounted for some of the variance in RTs, the ENSs no longer showed an overall effect of Word Frequency, but the SNSs did. The DNSs did not show a main effect of Word Frequency in either model.

The main effect of Word Ambiguity approached significance for the FNSs, with slower picture naming times for nouns with ambiguous endings. None of the three NNS language groups showed a main effect of Word Ambiguity; however, both the SNS and DNS participants, but not the ENSs, showed significant interactions involving Word Ambiguity in both the first and second models.[48]

Generally speaking, the SNSs showed faster RTs for unambiguous nouns and slower RTs for ambiguous nouns; however, this effect was not consistent across Word Frequency or Prime Compatibility conditions. The DNSs also showed faster RTs for unambiguous nouns, but only for low frequency nouns, whereas, conversely, RTs for high frequency exception nouns were faster. The complex interactions between Word Ambiguity, Prime Compatibility, and Word Frequency make it difficult to determine a pattern of facilitation or interference based on Word Ambiguity; however, the interactions do suggest that, like FNSs, the SNS and DNS participants are sensitive to noun endings in this task, whereas ENSs are not. To investigate whether this difference in

[48] Figures of these interactions are presented in Appendix K.

sensitivity was also evident in the offline gender assignment post-test, accuracy and RTs were compared for the ambiguous, unambiguous, and exception words that occurred in the gender priming task, as shown in Figures 19-20.

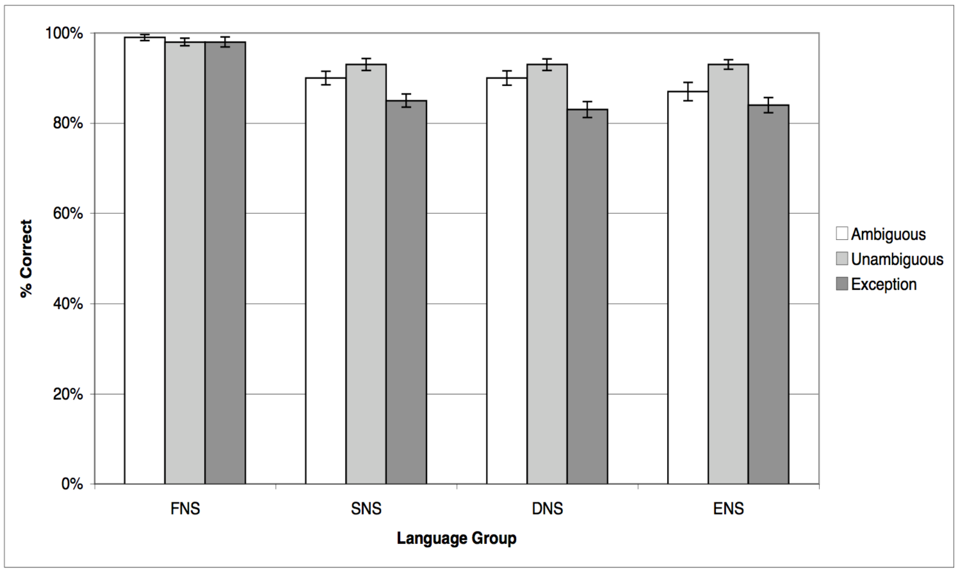


Figure 19. Accuracy (% correct) on ambiguous, unambiguous, and exception target nouns in gender assignment post-test for all four language groups

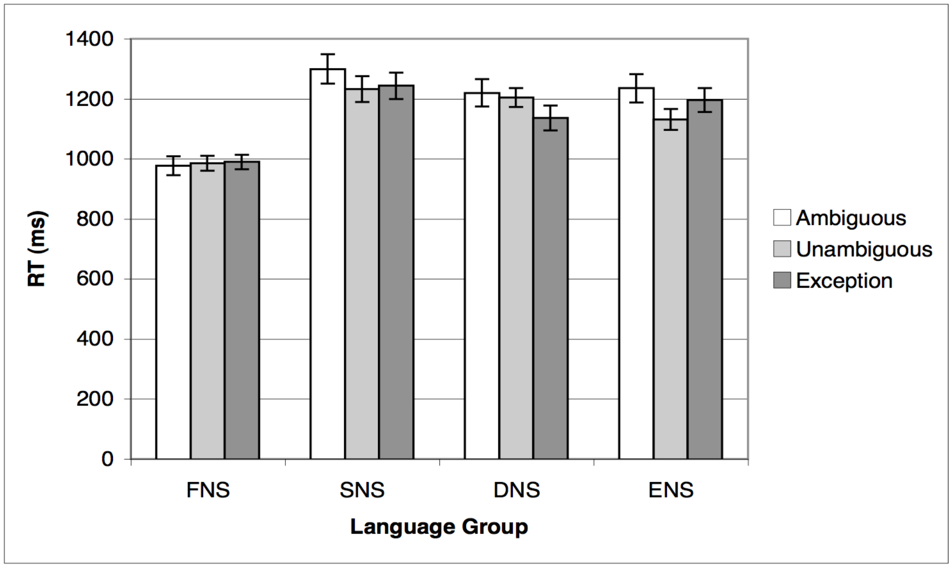


Figure 20. RT (ms) on ambiguous, unambiguous, and exception target nouns in gender assignment post-test for all four language groups

A series of paired samples t-tests (statistical results presented in Table 31) revealed no significant differences for FNSs for ambiguous, unambiguous, or exception words on either accuracy (most likely due to ceiling effects) or RT. The NNSs, however, all showed a significant difference between accuracy on unambiguous and exception words. The SNS and DNS participants also showed a significant difference in accuracy between ambiguous and exception words, and the DNS and ENS participants showed a significant difference for accuracy between ambiguous and unambiguous words. All three NNS groups showed a significant difference in RT between ambiguous and unambiguous words; the DNSs and ENSs, but not the SNSs, showed a significant difference in RT between unambiguous and exception words, and the SNSs showed a significant difference in RT between ambiguous and exception words.

Table 31

T-test Results for Gender Assignment (Accuracy and RT) on Ambiguous, Unambiguous, and Exception Target Nouns on Gender Priming Task

	FNS	SNS	DNS	ENS
Accuracy				
Ambiguous – Unambiguous	Not Sig.	Not Sig	$t(37) = -2.624$ $p < .05$	$t(41) = -3.846$ $p < .01$
Ambiguous – Exception	Not Sig.	$t(36) = -2.452$ $p < .05$	$t(37) = 4.25$ $p < .01$	Not Sig.
Unambiguous – Exception	Not Sig.	$t(36) = 5.142$ $p < .05$	$t(37) = 6.719$ $p < .01$	$t(41) = 5.648$ $p < .01$
RT				
Ambiguous – Unambiguous	Not Sig.	$t(36) = 2.545$ $p < .05$	$t(37) = 3.319$ $p < .01$	$t(41) = 3.782$, $p < .01$
Ambiguous – Exception	Not Sig.	$t(36) = 2.202$ $p < .05$	Not Sig.	Not Sig.
Unambiguous – Exception	Not Sig.	Not Sig.	$t(37) = -2.724$ $p < .01$	$t(41) = -3.636$ $p < .01$

Overall, the three NNS language groups are sensitive to phonological endings as an indicator of a noun's gender in an offline gender assignment context, as evident from highest accuracy on words with unambiguous endings, lower accuracy on words with ambiguous endings, and lowest accuracy on exception words, a finding that has been shown in previous studies (Hardison, 1992; Holmes & Dejean de la Bâtie, 1999). However, unlike for the FNSs, this sensitivity did not manifest itself at all for the ENSs in the gender priming task, and other than a weak pattern of facilitation in picture naming for low frequency unambiguous nouns, did not result in any clear pattern of facilitation or interference for the SNSs or DNSs.

As discussed above, the first HLM model was run for the NNS groups to determine whether they perform like FNSs on the gender priming task. A second HLM model including AO, number of years participant has known French, and number of years spent in France was run for the NNS groups. The purpose of this second analysis was to investigate whether, for some combination of these variables, an effect of congruency would emerge. However, no definitive congruency effects were found for any of the NNS groups. Another approach to determining whether gender priming occurs for some subset of participants is to consider the participants who did show priming effects and look at what these participants might have in common. The first step to this follow-up analysis was to set a standard of what qualified as priming effects. To this end, participants with a difference of 50ms[49] or greater between the incongruent and congruent conditions were considered to have shown priming effects (FNS = 8 [out of 21], SNS = 17 [out of 37], DNS = 17 [out of 38], ENS = 18 [out of 42]).

First, the distribution among Counterbalancing Groups was compared for those who showed priming effects to determine whether priming effects were due to list effects. The distribution is presented in Table 32.

[49] The difference of 50ms was selected based on Alario et al. (2004), who found a difference in RTs of approximately 50ms between the incongruent and congruent conditions.

Table 32

Distribution of Participants Who Show Priming Effects Among Group

	Group A	Group B	Group C
FNS	2	2	4
SNS	4	0	12
DNS	4	4	9
ENS	6	3	8
Total	16	9	35

Group C had the greatest number of participants who showed priming effects, and it was also Group C who saw List 3 in the congruent condition. As discussed in Appendix K, RTs for List 3 were always fastest, regardless of congruency condition, for the NNS participants. Although list effects were not found for the FNS participants, the over-representation of Group C among the participants who showed priming may be explained, at least in part, by list effects.

Next, a series of independent sample t-tests were run, separately for each language group, with priming as the grouping variable (those who showed priming effects vs. those who did not), and a number of test variables selected based on their potential relevance to gender priming: gender assignment accuracy and RT (on words that occurred in gender priming task only), overall grammaticality judgment task accuracy, grammaticality judgment filler sentence accuracy, grammaticality judgment accuracy on the close condition, grammaticality judgment accuracy on the far condition, and for the NNSs, AO, number of years spent in France, and number of years participant has known French. Gender assignment accuracy and grammaticality judgment overall and filler sentence accuracy may be considered as general proficiency measures. That is, perhaps only participants who achieved high accuracy on these tasks are also the ones showing gender priming effects. Because accuracy on the grammaticality judgment far condition requires maintaining the gender of the target noun throughout the sentence, it is possible that the same process used to maintain gender activation also relates to gender priming. Therefore, participants who achieved high accuracy in the grammaticality judgment far condition may be the same participants who showed priming effects. And finally, the language experience factors may be related to which NNSs showed priming effects, even though these factors did not emerge as significant in the HLM analysis.

No significant difference between participants who showed priming effects and participants who did not were found for any of the variables for any language group. That is, none of these variables stand out as a common factor among participants who show (or do not show) priming effects. Finally, data from the correlation matrix were examined to determine whether WM capacity plays a role in gender priming. However, there was no correlation between the difference between incongruent and congruent RTs and absolute span scores for any of the language groups. A few spurious correlations emerged between the incongruent-congruent RT effect and other variables (i.e., higher achievement on grammaticality judgment correct filler sentences correlated to a smaller priming effect for SNSs, $r(37) = -.331$, $p <$ 05, but a larger priming effect for DNSs $r(38) = .339$, $p < .05$); however, these conflicting correlations are difficult to interpret and do not represent a pattern within or among language groups. On the whole, after considering a number of potential variables, from language experience factors to performance on other tasks, it can be concluded that no common variable emerged among participants who show priming effects, nor did any variable stand out as a potential predictor for gender priming effects.

8.4.6 Conclusion

Overall, the HLM analysis revealed clear gender priming effects for FNSs, but there was no strong evidence for gender priming for any of the NNS participants. That is, the FNS participants were faster at naming target pictures when the prime's gender was congruent with the target noun as compared to when the primes' gender was incongruent with the target noun, but the NNSs did not show a consistent difference in RTs based on prime-target gender congruency. Even when item characteristics were taken into account – specifically Word Ambiguity and Word Frequency – there was no set of items for which NNS priming consistently occurred. In other words, not only was there no definitive main effect of gender priming, but there was no evidence of gender priming for any subset of items, such as high frequency words, or words with unambiguous noun endings. The same was true for participant characteristics: gender priming did not occur for any subset of participants based on the number of years a participant had known French, the number of years spent in France, or performance on other tasks in this study. Although the DNSs with the earliest AO showed some evidence of priming effects, the lack of priming effects for the DNS group as a whole, despite their overall early AO, suggests that AO is not a predictive factor.

To conclude, only the third hypothesis for this task was supported. The hypotheses are repeated below for the reader's convenience.

1a. Spanish learners of French will show evidence of a gender storage and nodal system.

1b. Dutch learners of French will show evidence of a gender storage and nodal system.

1c. English learners of French will not show evidence of a gender storage and nodal system.

Contrary to expectations, none of the NNS language groups showed clear evidence of a gender storage and nodal system, as indicated by the inconsistent priming effects in the gender priming task. Regardless of L1-L2 gender-system similarity, the learners of French did not perform similarly to the FNSs, thus, demonstrating that, even at an advanced level of proficiency, they do not represent grammatical gender similarly to FNSs, and that L1 similarity is not a facilitating factor.

8.5 Grammaticality Judgment Task

8.5.1 French Native Speakers

After completing the grammaticality judgment task, each FNS participant was asked to review the task sentences offline. Due to the RSVP paradigm used in this task, it is common for participants to have a certain percent of unintentional button presses (approximately 2% in the pilot), that is, button presses that do not correspond with the intended judgment of a sentence. Therefore, rather than relying solely on the participants' online judgments, the FNSs were asked to complete an offline task in which they reviewed the sentences for which their judgments during the online task differed from those of the researcher. This review process allowed the participants to identify any unintentional button presses, resulting in a more accurate tally of NS agreement on the task sentences. After a participant completed the online task, the researcher pulled up a spreadsheet displaying all 144 sentences to review with the participant. The spreadsheet was programmed to automatically highlight the sentences for which the participant's online judgment did not match that of the researcher. The researcher asked the participant to read each highlighted sentence and determine whether he/she thought the sentence was grammatically correct or incorrect. The participant's response was recorded on the spreadsheet. In addition, if the participant thought the sentence was incorrect, he/she was asked to identify the error.

Pre-determined criteria dictated that all sentences for which more than one of the 21 FNS participants disagreed with the researcher's judgment would be removed from all subsequent analyses. No target sentences were removed, as there were only three sentences for which one of the 21 NSs did not identify the gender agreement error during the offline review, and none for which more than one NS did not identify the error. Furthermore, it was the same FNS who accepted the three sentences as correct, despite the noun-adjective gender agreement error that the remaining 20 FNSs identified as incorrect. The remaining 45 target sentences had 100% FNS agreement. There were six (out of 96) filler sentences for which one FNS participant did not agree with the researcher's judgment. One was an incorrect filler that one participant thought to be correct, and five were correct filler sentences which were deemed incorrect due to stylistic preferences. These six sentences were not removed from the task because the remaining 20 FNSs agreed with the researcher's judgment of these sentences. In addition, there were four correct filler sentences that either two or three FNSs judged to be incorrect. These four sentences were removed from all subsequent analyses. The remaining 86 filler sentences had 100% agreement. Overall, only four of the 144 sentences were removed due to more than one FNS disagreeing with the researcher's judgment of the sentences' accuracy. The remaining 140 sentences were included in the FNS and NNS analyses.

Prior to carrying out the accuracy analysis, sentences for which a participant responded too early were removed for that participant. For incorrect sentences, a response was considered too early if the participant made an accurate judgment *before* the word containing the error appeared on the screen. Because it is not possible to judge a sentence based on the intended error before that error is presented, these judgments were either the result of an unintentional button press or based on an error other than the one intended by the researcher. For correct sentences, a response was considered too early if the participant made an accurate judgment before the final word of the sentence appeared on the screen. In the case of early responses on correct sentences, the participant may have predicted the sentence was coming to an end and correctly assumed that no error would appear; however, in order to remain consistent, these trials were also removed from the analyses. Consequently, out of a total of 2940 trials, 59 (2.0%; 30 correct filler, 5 incorrect filler, and 24 target) were removed from the accuracy analysis. Finally, an arcsine transformation was calculated for the accuracy data in order to meet the homogeneity of variance assumption across subjects.

The RT analysis was conducted only on sentences that were accurately judged during the online task, as standard in the literature. To this end, of 2940 trials, 152 trials (5.2%) were removed due to inaccurate responses. RTs represent the point at which the participant pushed a button on the button box indicating whether the sentence was correct or incorrect. The RT for the incorrect sentences was measured from the onset of the word containing the error, that is, the word at which point the sentence became ungrammatical, and the RT for correct sentences was measured from onset of the final word of the sentence. As described above, if a participant judged a sentence before the presentation of the word containing the error or the final word of a correct sentence, a negative RT was recorded; these early response trials (59 trials, 2.0%) were removed first. In addition, one trial was removed due to a technical difficulty in which the Psyscope program was paused and RT data from this sentence were lost. Outliers 2.5 standard deviations above or below the mean were calculated separately for the target sentences, incorrect fillers, and correct fillers; to this end, a total of 67 trials (2.3%) were removed. To summarize, 279 trials out of a total of 2940 (9.5%) were removed due to inaccurate judgments, negative RTs, technical difficulty, and RT outliers. The remaining 2661 trials were included in the RT analysis.

Before proceeding with the accuracy and RT analyses, the reliability, or internal consistency, of the target sentences was calculated to determine how well the items measure a unidimensional construct, that is, whether the items within a condition produce similar scores within participants. Because the target sentences all measure noun-adjective gender agreement, they were intended to represent the same construct. Cronbach's alpha for the target sentences was .816 (48 items), indicating high internal consistency and confirming that the target sentences represent the same construct.

After cleansing the data and confirming target sentence reliability, accuracy and RTs were compared across the target, incorrect filler and correct filler sentences. FNS mean accuracies and RTs for these three types of sentences are displayed in Figures 21 and 22. A paired-samples t-test for accuracy showed no significant difference between the target and filler (both correct and incorrect) sentences. However, a significant difference was found between both the incorrect filler and correct filler sentences, $t\ (20) = -7.722$, $p < .01$, and the incorrect filler and target sentences, $t\ (20) = 3.586$, $p < .01$. No significant difference was found between the target and correct filler sentences. Overall, the FNSs performed well on both the target and filler sentences, with the incorrect filler sentences being the most difficult. Performance on these sentences will be explored further in Section 8.5.3 using a d' analysis. The FNSs were also fastest at judging the correct filler sentences, and slowest at judging the incorrect filler sentences. A paired-samples t-test for RT showed significant differences between target and correct filler sentences, t (20) = 3.594, $p < .01$, correct fillers and incorrect filler sentences, t (20) = 5.023, $p < .01$, and target and incorrect filler sentences, t (20) = -3.400, $p < .01$. These results are consistent with the accuracy results in that the FNSs were slower to judge the sentences on which they achieved lower accuracy, indicating there was no speed-accuracy trade-off.

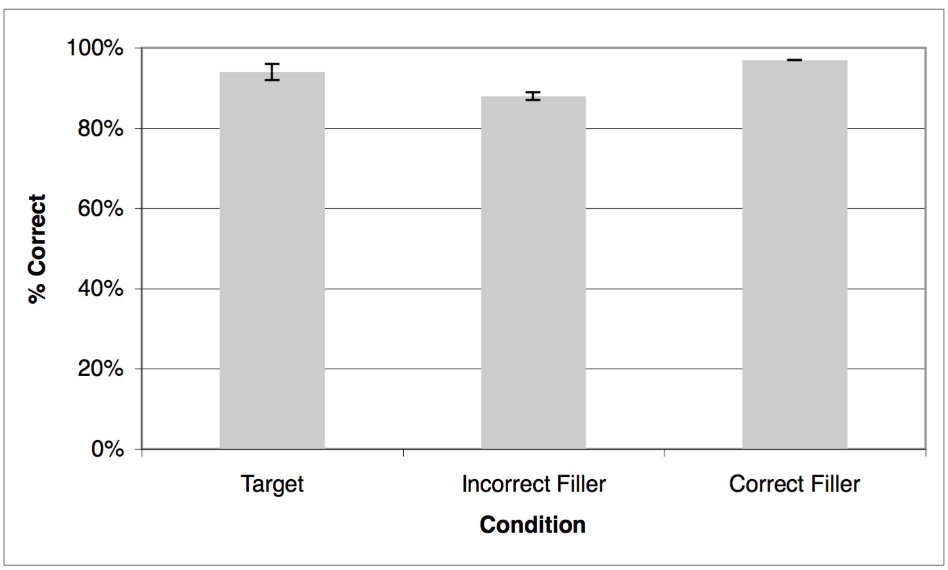


Figure 21. FNS accuracy (%) on target, incorrect filler, and correct filler sentences

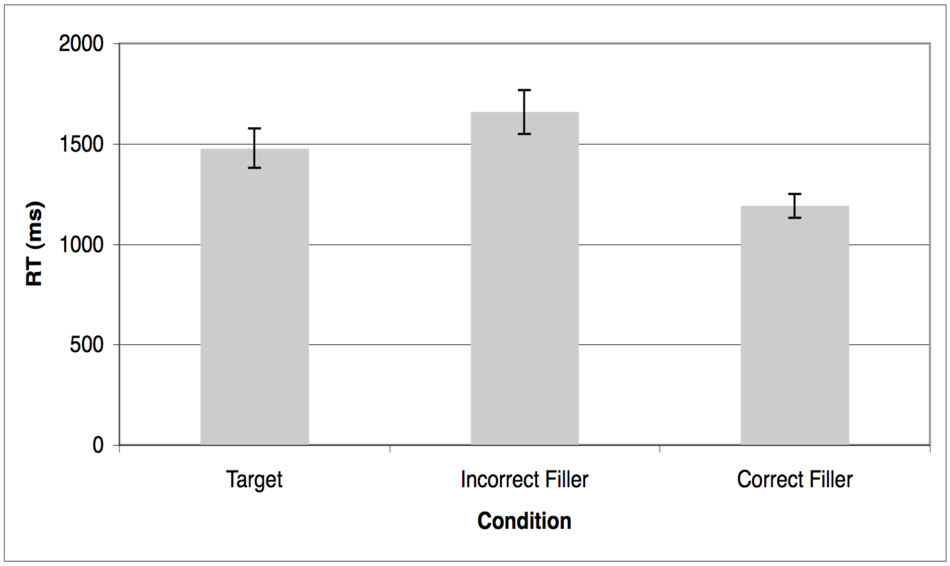


Figure 22. FNS RT (ms) on target, incorrect filler, and correct filler sentences

Within the target sentences, in order to determine in which conditions the participants achieved the highest accuracy and fastest RTs, percent accuracy scores and RTs by subject ($F1$) and item ($F2$) were entered into a one-way repeated-measures analysis of variance (ANOVA)[50] with noun-adjective distance (Distance), gender cue (Cue), and target noun gender (Gender) as the independent variables.

For accuracy, there was a main effect of Gender for both the subject, $F1$ (1, 20) = 11.742, $p < .01$, and item, $F2$ (1, 40) = 6.752, $p < .05$, analyses. FNSs achieved higher accuracy (M = 96%, SD = 6%) when judging sentences in which a masculine noun was modified by a feminine adjective than on sentences in which a feminine noun was modified by a masculine adjective (M = 92%, SD = 10%). There were no other main effects or interactions that were significant for both subject and item analyses; however,

[50] Multilevel modeling (specifically, HLM) was deemed appropriate for the gender priming task analysis. Due to the nature of the grammaticality judgment task, which does not suffer as much from missing data, the main argument for using multilevel modeling for the gender priming task does not apply. Therefore, repeated- measures ANOVAs were used in the grammaticality judgment task analysis.

there was a significant interaction for RT between Cue and Gender in the subject analysis, F (1, 20) = 12.430, $p < .01$. When a gender cue was provided, FNSs were faster to judge sentences in which the target noun was masculine, but when no gender cue was provided, the NSs were equally fast at judging sentences regardless of the target noun's gender, as shown in Figure 23.

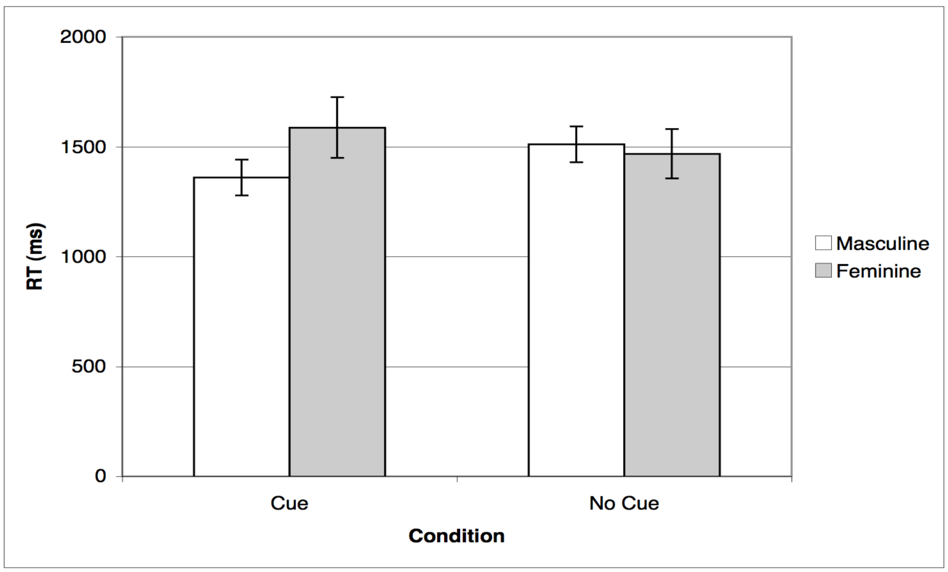


Figure 23. FNS RT (ms) on sentences in Cue and Gender conditions

Finally, when low and high span groups were included in the item analysis[51] as the within-items variable, a significant main effect of Span appeared for both accuracy, F (1, 40) = 7.722, $p < .01$, and RT, F (1, 40) = 6.49, $p < .01$. FNS high span participants achieved higher accuracy (M = 97%, SD = 9%) and faster RTs (M = 1550, SD = 382) on the target sentences than the low span participants (accuracy: M = 91%, SD = 12%; RT:

[51] WM span was not considered in the subject analysis because including WM span as a between-subjects variable would have created a power issue due to the number of participants removed in order to create two distinct span groups. In addition to including span group as variable in the item analysis, WM span will be considered as a continuous variable in the correlation analysis for all four language groups.

$M = 1761$, $SD = 421$). These differences between span groups suggest that WM capacity plays a role in a participant's overall ability to detect gender errors, although it does not interact with any of the main variables. To determine whether span is specifically related to the ability to detect gender errors, or more generally to performance on this grammaticality judgment task, item analyses for filler sentence accuracy and RT, with span as the within-items variable, were conducted.

No effect of span was found for accuracy, even when correct and incorrect fillers were considered independently. This finding suggests that span does indeed play a role for target sentence accuracy, especially given that an effect of span was found for target sentences despite near-ceiling accuracy. However, it should be noted that neither the correct nor incorrect fillers measure a unidimensional construct, as determined by their low internal consistency (Cronbach's alpha for correct fillers: -.239 [68 items]; Cronbach's alpha for incorrect fillers: .076 [24 items]). In other words, had the incorrect filler sentences measured the same construct, for example, negation, it is possible that an effect of span for accuracy would have emerged. Finally, an effect of span was found for RT on all fillers combined, $F\,(1, 91) = 7.248$, $p < .01$, and incorrect fillers, $F\,(1, 23) = 4.354$, $p < .05$, and approached significance for correct fillers, $F\,(1, 67) = 3.728$, $p = .058$, with faster RTs for the high span group as compared to the low span group. The faster RTs for the high span group on the filler sentences suggest span is related to an ability to detect errors on the task in general. The role of span in gender agreement ability will be considered further in a correlation matrix presented in Section 8.5.3.

Overall, FNSs performed well on all conditions of the grammaticality judgment task. Mean accuracies for each condition are displayed in Figure 24, and were used as a baseline to compare to the NNS groups.

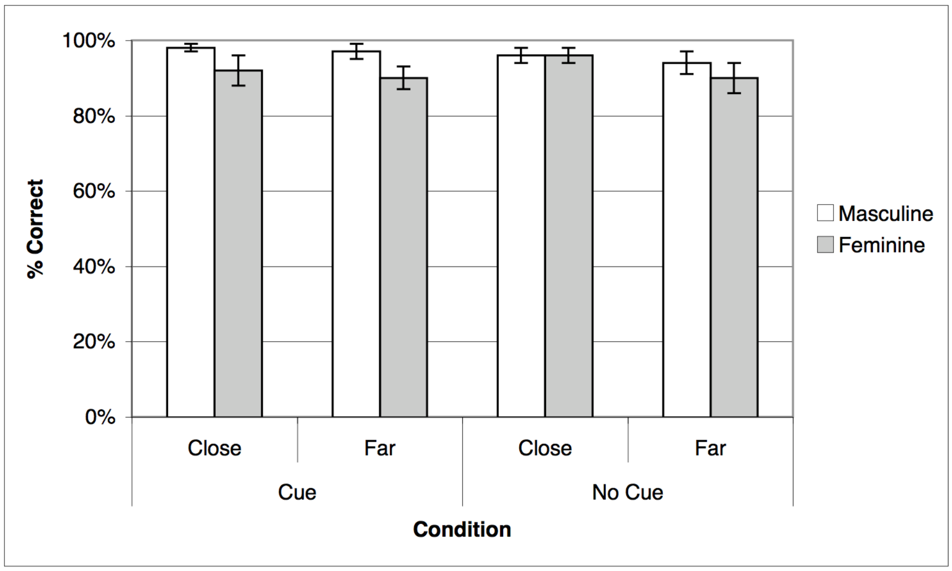


Figure 24. FNS accuracy (%) on target sentence conditions

8.5.2 Non-native Speakers

For all three NNS language groups, the same data cleansing process and reliability, accuracy and RT analyses were carried out as for the FNSs. First, the four correct filler sentences that either two or three FNS participants judged to be incorrect in the post-task review process, as discussed above, were removed from all NNS analyses. Second, reliability, or internal consistency, for target sentences was calculated to determine whether the items measure a unidimensional construct for each language group. Third, trials on which a participant made an early response, that is, an accurate judgment before the word containing the error appeared on the screen, or in the case of correct sentences, before the final word of the sentence appeared on the screen, were removed from the accuracy analysis. Fourth, an arcsine transformation was calculated for the accuracy data. Finally, before carrying out the RT analyses, all trials with early responses, inaccurate judgments, and RT outliers were removed. The data removal process resulted in several empty RT cells across all combinations of the three conditions (Distance, Cue, Gender) for some individuals in each NNS language group; these empty cells were replaced with the serial mean for that condition. For example, an empty RT cell for the close-cue-masculine condition for a given participant was replaced with the mean RT of the remaining participants (within the same language group) for that condition. The internal reliability, number of trials removed, and number of empty RT cells that were replaced with serial means will be reported for each language group before presentation of the results.

In addition, the same analyses that were conducted for the FNSs were conducted for all three NNS groups, starting with a series of paired-samples t-tests to compare accuracy and RT performance across the target, incorrect filler and correct filler sentences. Next, both accuracy (%) and RT (ms) by subject ($F1$) and item ($F2$) were submitted to one-way repeated-measures ANOVAs for the target sentences, with Distance, Cue, and Gender as the independent variables. Finally, low and high WM span group were included in the item analysis ($F2$) as the within-items variable to determine the role of WM capacity in accuracy and RT performance.

8.5.2.1 Spanish Native Speakers

The reliability was calculated for the target sentences; Cronbach's alpha was .929 (48 items), indicating high internal consistency. Prior to conducting the accuracy analysis, out of a total of 5180 trials, 72 accurately judged sentences (32 correct filler, 6 incorrect filler, and 34 target) were removed due to early responses. Before conducting the RT analysis, in addition to removing the 72 early response trials, 1348 trials (26.0%) were removed due to inaccurate responses, and 99 trials (1.9%) were removed as outliers. Overall, 1519 trials out of a total of 5180 (29.3%) were removed from the RT analysis due to early responses, inaccurate judgments, and outliers, with 3661 trials remaining. A total of 27 empty cells (9% of 296 cells) across 37 participants and all combinations of the three conditions (distance, cue, gender) were replaced with the serial mean for that condition.

A series of paired-samples t-tests for accuracy and RT showed significant differences between all sentence types. SNS participants performed better on filler sentences than target sentences, t (36) = -10.586, $p < .01$, with significantly higher accuracy on both correct, t (36) = -10.850, $p < .01$, and incorrect, t (36) = -6.791, $p < .01$, fillers than on target sentences, and higher accuracy on correct filler sentences than incorrect filler sentences, t (36) = -6.482, $p < .01$. The differences in RTs mirrored those for accuracy, with faster RTs on filler sentences than target sentences, t (36) = 6.761, $p < .01$, faster RTs on both correct, t (36) = 7.064, $p < .01$, and incorrect, t (36) = 4.087, $p < .01$ fillers as compared to target sentences, and finally, faster RTs on correct fillers than on incorrect fillers, t (36) = 5.218, $p < .01$. SNS mean accuracies and RTs for the target, incorrect, and correct fillers are displayed in Figures 25 and 26.

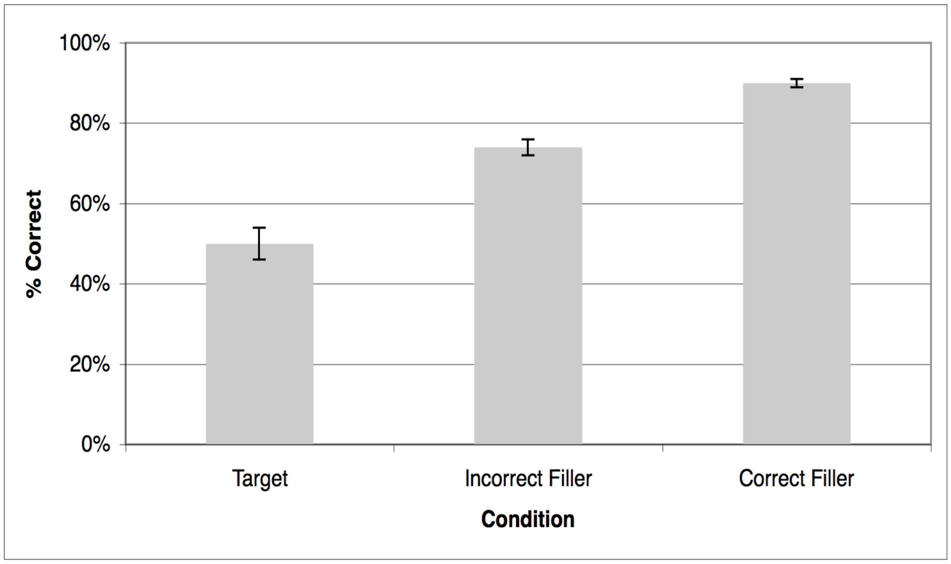


Figure 25. SNS accuracy (%) on target, incorrect filler, and correct filler sentences

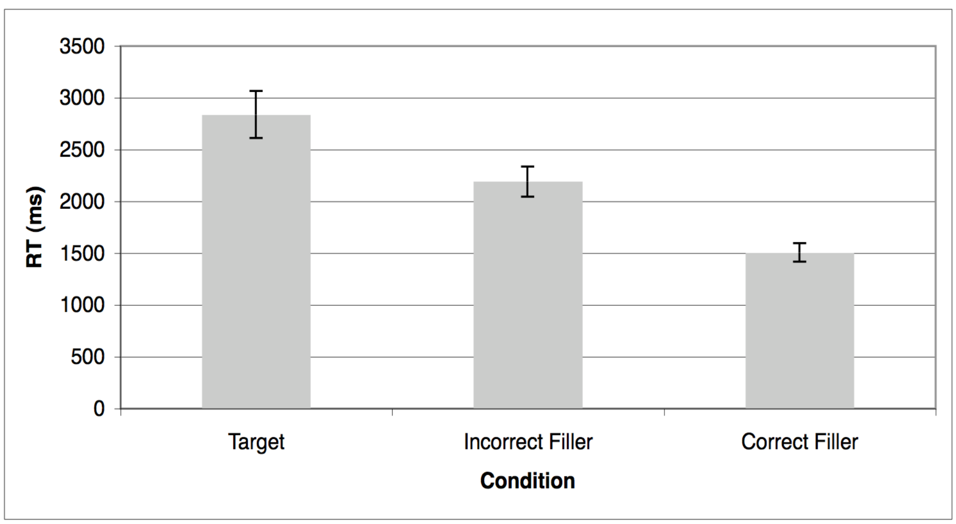


Figure 26. SNS RT (ms) on target, incorrect filler, and correct filler sentences

Overall, the SNS participants achieved the highest accuracy and the fastest RTs on the correct filler sentences and the lowest accuracy and slowest RTs on the target sentences. As seen with the FNSs, the SNSs were faster to judge the types of sentences on which they achieved higher accuracy, and slower to judge the types of sentences on which they achieved lower accuracy, indicating that there was no speed-accuracy trade-off.

The repeated-measures ANOVAs for accuracy revealed a main effect of Gender in both the subject *F*1 (1, 36) = 31.639, $p < .01$, and item, *F*2 (1, 40) = 9.586, $p < .01$ analyses. SNSs achieved higher accuracy (*M* = 57%, *SD* = 4%) on sentences in which a feminine noun was modified by a masculine adjective than on sentences in which a masculine noun was modified by a feminine adjective (*M* = 43%, *SD* = 4%). This pattern is consistent with the main effect of Gender RT for both the subject *F* (1, 36) = 23.4, $p < .01$, and item, *F* (1, 40) = 7.068, $p < .05$ analyses; that is, the sentences that were more difficult for the SNSs also had slower RTs (*M* = 3113, *SD* = 257) than the sentences that were easier (*M* = 2617, *SD* = 210). Interestingly, this is the opposite pattern of that seen with the FNSs, who performed better on sentences with a masculine noun modified by a feminine adjective.

There were no other main effects or interactions that were significant for both subject and item analyses; however, for accuracy there was a significant main effect in the subject analysis for Cue, *F*1 (1, 36) = 10.387, $p < .01$. SNSs achieved higher accuracy on sentences with a gender cue (*M* = 54%, *SD* = 4%) than sentences with no gender cue (*M* = 46%, *SD* = 4%).

Finally, when low and high span groups were included in the item analysis as the within-items variable, a significant main effect of Span appeared for accuracy, *F* (1, 40) = 11.410, $p < .01$. SNS high span participants achieved higher overall accuracy on the target sentences (*M* = 56%, *SD* = 20%) than the low span participants (*M* = 49%, *SD* = 19%); however, the low accuracy for both the low and high span participants indicates that a high span does not provide an impressive advantage. Although a main effect of Span was not found for RT, there was a significant interaction for RT between Span and Gender, F (1, 40) = 5.512, $p < .05$, with high span participants performing faster than low span participants on sentences with feminine target nouns, but slower than low span participants on sentences with masculine target nouns (Figure 27). This finding suggests that high span participants are more sensitive to gender.

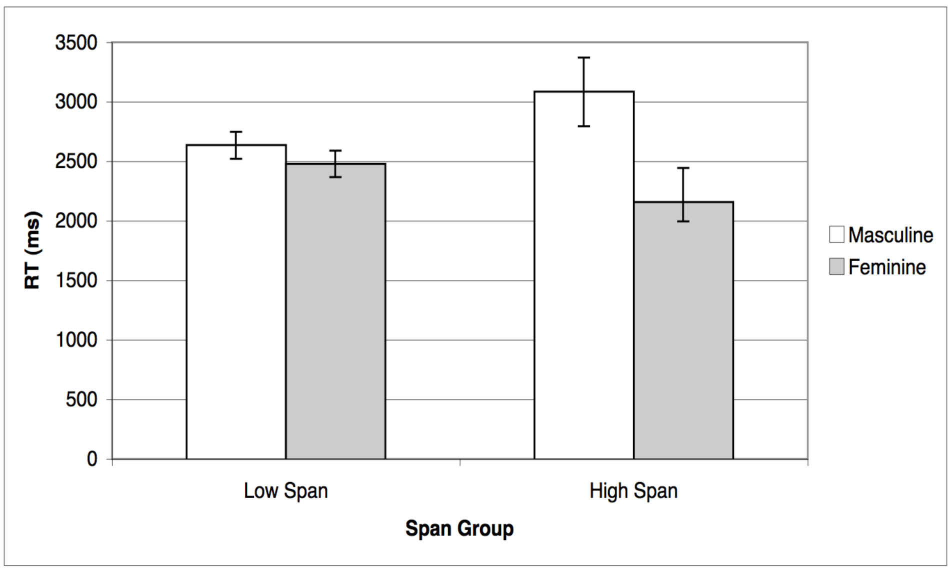


Figure 27. SNS RT (ms) on sentences in Gender condition for low and high span groups

An item analysis for filler sentence accuracy and RT, with span as the within-items variable, revealed no effect of span, even when the correct and incorrect fillers were considered independently, suggesting that span is specifically related to the ability to detect gender errors.

Overall, the SNSs performed poorly on the target sentences, with at chance accuracy scores. Mean accuracies for each condition are displayed in Figure 28.

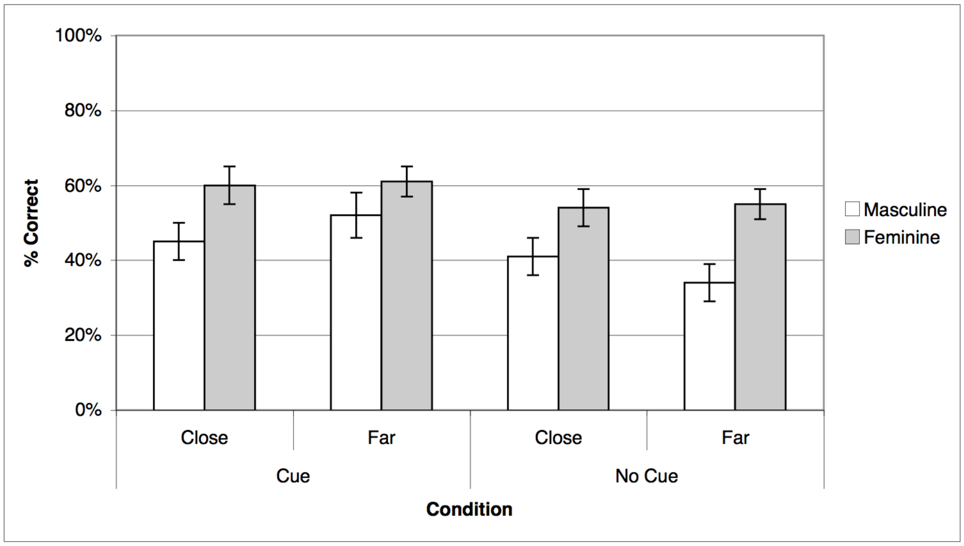


Figure 28. SNS accuracy (%) on target sentence conditions

Furthermore, in contrast with the FNSs who achieved higher accuracy on the target sentences than on the incorrect filler sentences, the SNSs achieved higher accuracy on the incorrect filler sentences than on the target sentences. This pattern confirms previous research that noun-adjective gender agreement is indeed difficult for even advanced learners of French, and suggests that L1-L2 gender-system similarity does not facilitate realization of L2 gender agreement.

8.5.2.2 Dutch Native Speakers

The reliability for the target sentences indicates high internal consistency (Cronbach's alpha = .920, 48 items). Prior to conducting the accuracy analysis, out of a total of 5320 trials, 93 accurately judged sentences (2%; 38 correct filler, 6 incorrect filler, and 49 target) were removed due to early responses. Before conducting the RT analysis, in addition to the 93 early response trials, 1055 trials (19.8%) with inaccurate responses and

114 outlier trials (2.1%) were removed. Overall, 1262 trials out of a total of 5320 (23.7%) were removed from the RT analysis due to early responses, inaccurate judgments, and outliers, leaving 4058 remaining trials. A total of 19 empty RT cells (6% of 304 cells) across 38 participants and all combinations of the three conditions (Distance, Cue, Gender) were replaced with the serial mean for that condition.

The series of paired-samples t-tests for accuracy and RT showed significant differences for all sentence types. DNS participants performed significantly better on filler sentences than target sentences, $t\ (37) = -11.504$, $p < .01$, with significantly higher accuracy on both correct, $t\ (37) = -11.441$, $p < .01$, and incorrect, $t\ (37) = -8.559$, $p < .01$, fillers than on target sentences, and higher accuracy on correct filler sentences than incorrect filler sentences, $t\ (37) = -4.412$, $p < .01$. The DNSs were also faster at judging sentences on which they achieved higher performance, with faster RTs on filler sentences than target sentences, $t\ (37) = 7.307$, $p < .01$, faster RTs on both correct, $t\ (37) = 7.714$, $p < .01$, and incorrect, $t\ (37) = 3.297$, $p < .01$ fillers than on target sentences, and faster RTs on correct fillers than on incorrect fillers, $t\ (37) = 4.362$, $p < .01$. This accuracy and RT pattern is similar to that seen for the SNSs, and is displayed in Figures 29 and 30.

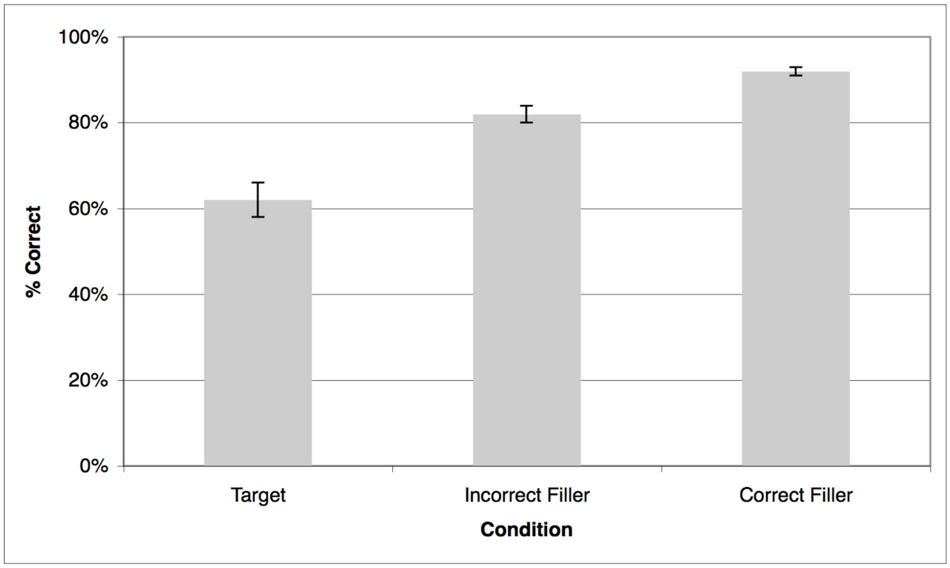


Figure 29. DNS accuracy (%) on target, incorrect filler, and correct filler sentences

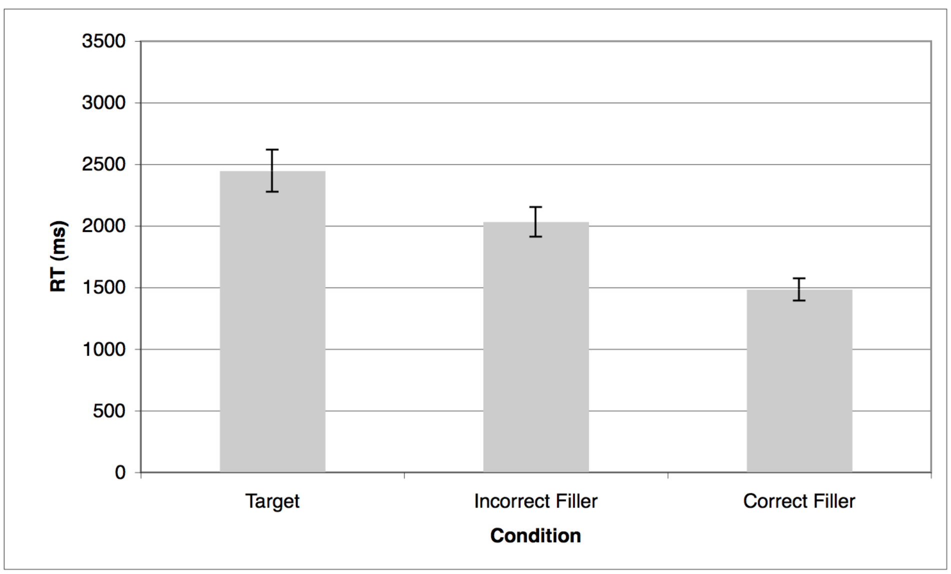


Figure 30. DNS RT (ms) on target, incorrect filler, and correct filler sentences

The subject ($F1$) and item ($F2$) ANOVAs revealed main effects of both Cue, $F1\ (1, 37) = 69.363$, $p < .01$, $F2\ (1, 40) = 25.358$, $p < .01$, and Gender, $F1\ (1, 37) = 9.049$, $p < .01$, $F2\ (1, 40) = 9.080$, $p < .01$, as well as a three-way interaction for Distance, Cue, and

Gender, *F1* (1, 37) = 9.6, $p < .01$, *F2* (1, 40) = 5.256, $p < .05$. As shown in Figure 31, in the close condition, DNSs achieved higher accuracy on sentences in which a feminine noun is modified by a masculine adjective, regardless of whether a gender cue is provided. However, in the far condition, DNSs performed slightly better on sentences in which a masculine noun is modified by a feminine adjective when a cue is given, but performed better on sentences in which a feminine noun is modified by a masculine adjective when no cue is given.

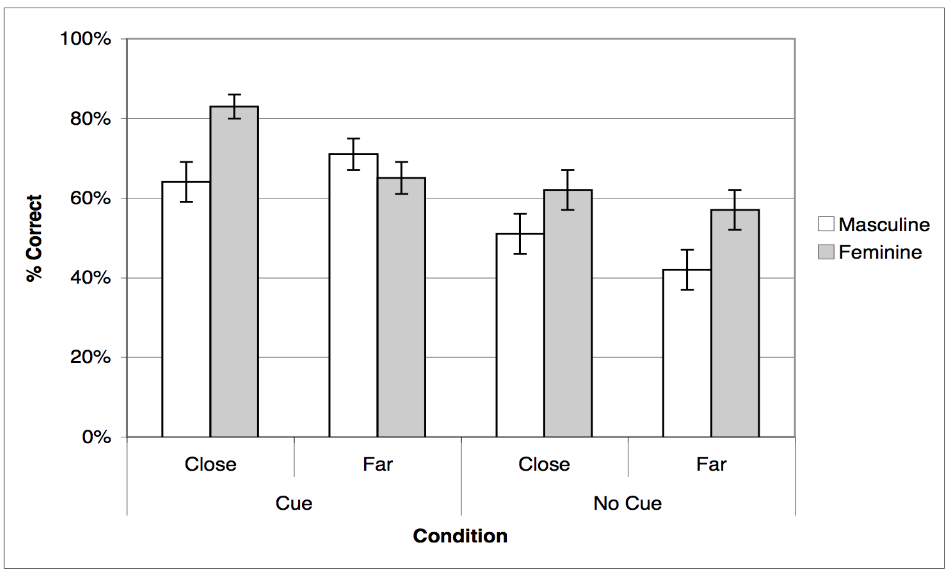


Figure 31. DNS accuracy (%) on target sentence conditions

There were no other main effects or interactions for accuracy that were significant in both the subject and item analyses; however, in the subject analysis, a main effect of Distance, *F*1 (1, 37) = 13.036, $p < .01$, and a two-way interaction between Distance and Gender, *F*1 (1, 37) = 6.975, $p < .05$, were significant.

For RT, only a main effect of Cue was found in both subject and item analyses, $F1$ (1, 37) = 8.069, $p < .01$, $F2(1, 40) = 5.376$, $p < .05$, with participants judging sentences with a gender cue faster ($M = 2348$, $SD = 192$) than sentences with no gender cue ($M = 2538$, $SD = 168$). In the subject analysis only, there were main effects of Distance, $F1$ (1, 37) = 4.601, $p < .05$, and Gender, F (1, 37) = 33.4, $p < .01$, as well as a significant two-way interaction between Cue and Gender, F (1, 37) = 7.906, $p < .01$. The DNS participants were slightly faster ($M = 2346$, $SD = 202$) at judging sentences in which the adjective occurred directly after the target noun, as compared to sentences in which the adjective occurred several words after the target noun ($M = 2556$, $SD = 162$). In addition, DNSs judged sentences with a feminine noun modified by a masculine adjective faster than sentences with a masculine noun modified by a feminine adjective when no cue was provided, but RTs were similar on sentences with masculine and feminine nouns when a cue was provided, as shown in Figure 32. The RTs in this interaction are consistent with the general trend of higher accuracy on sentences with a gender cue as compared to sentences with no gender cue, and sentences with a feminine target noun as compared to sentences with a masculine target noun.

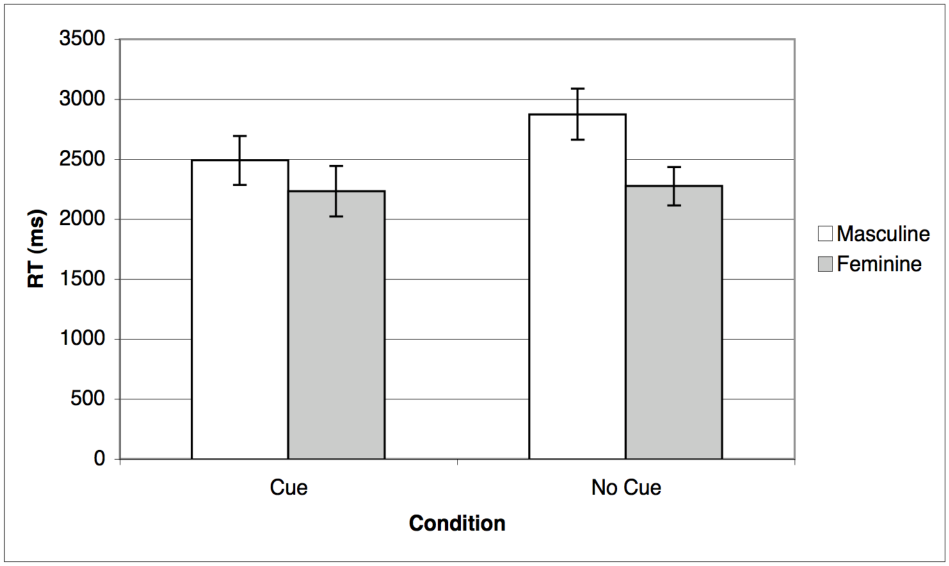


Figure 32. DNS RT (ms) on sentences in Cue and Gender conditions

Finally, the item analysis with low and high span group as the within-items variable showed a main effect of Span for accuracy F (1, 40) = 8.456, $p < .01$, though not for RT. However, Span did interact with Distance and Gender RT, F (1, 40) = 5.242, $p < .05$. Participants in the high span group achieved higher overall target sentence accuracy (M = 66%, SD = 19%) than participants in the low span group (M = 58%, SD = 20%). Although both low and high span groups showed similar RTs for sentences with masculine and feminine target nouns in the far condition, the high span group showed faster RTs on sentences with a feminine noun as compared to sentences with a masculine noun in the close condition, whereas the low span group did not show this difference. The RTs for the low and high span groups Distance and Gender are presented in Figure 33. Faster RTs for high span participants on sentences with feminine nouns occurred for SNSs as well; however the interaction with Distance did not. It may be that a processing advantage for sentences with feminine nouns modified by masculine adjectives only occurs in the close condition for the DNSs, whereas it occurred regardless of noun-adjective distance for the SNSs.

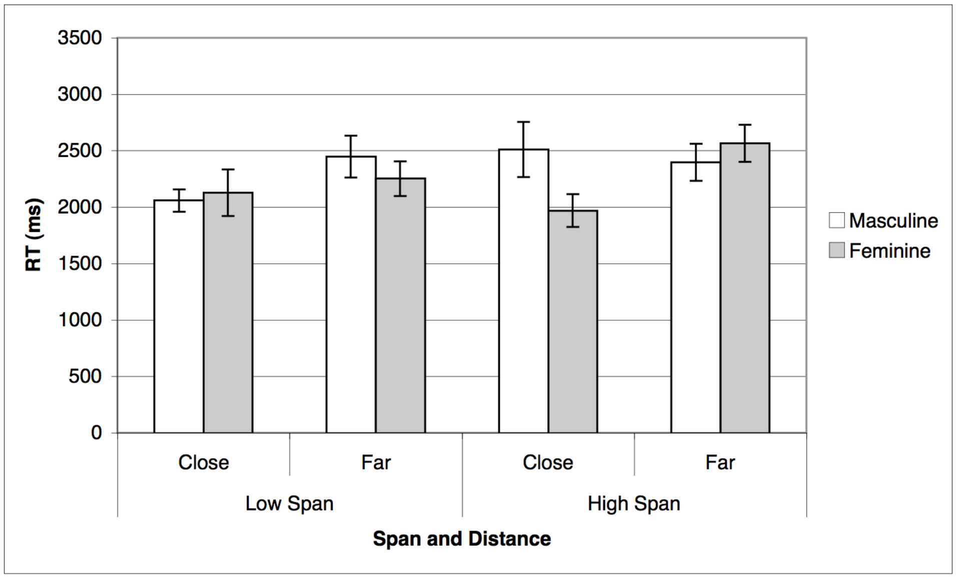


Figure 33. DNS RT (ms) on sentences in Cue and Gender conditions for low and high span groups

An item analysis for filler sentence accuracy and RT, with span as the within-items variable, was also conducted. A main effect of span for accuracy on all fillers combined, F (1, 91) = 9.155, $p < .01$, and correct fillers, F (1, 67) = 17.894, $p < .01$, indicated that low span participants performed better then high span participants. There was no main effect of span on incorrect fillers. Furthermore, a main effect of RT indicated that low span participants were faster than high span participants on correct fillers, F (1, 67) = 48.208, $p < .01$, but slower on incorrect fillers, F (1, 23) = 11.004, $p < .03$. No main effect of RT was found for all fillers combined. Perhaps this reverse pattern in RT for low and high span participants indicates a difference in strategy, with the low span participants responding quickly at the end of a sentence if they hadn't noticed an error, and high span participants taking more time to be sure no error was present.

Overall, similar to the SNS participants, the DNSs did not achieve native-like accuracy on the grammaticality judgment task sentences. Furthermore, the target sentences were the most difficult and the correct filler sentences were the least difficult.

8.5.2.3 English Native Speakers

The reliability for the target sentences indicates high internal consistency (Cronbach's alpha = .938, 48 items). Prior to conducting the accuracy analysis, out of a total of 5880 trials, 69 accurately judged sentences (1%; 24 correct filler, 6 incorrect filler, and 39 target) were removed due to early responses. Before conducting the RT analysis, in addition to the 69 early response trials, 1468 trials (25.0%) were removed due to inaccurate responses and 125 trials (2.1%) were removed as outliers. Overall, 1662 trials out of a total of 5880 (28.3%) were removed from the RT analysis due to early responses, inaccurate judgments, and outliers, with 4218 trials remaining. A total of 28 empty cells (8% of 336 cells) across 42 participants and all combinations of the three conditions (Distance, Cue, Gender) were replaced with the serial mean for that condition.

The paired-samples t-tests for both accuracy and RT showed significant differences among all sentence types. ENS participants performed significantly better on filler sentences than target sentences, t (41) = -11.174, $p < .01$, with significantly higher accuracy on both correct, t (41) = -10.928, $p < .01$, and incorrect, t (41) = -6.429, $p < .01$, fillers than on target sentences, and higher accuracy on correct filler sentences than incorrect filler sentences, t (41) = -5.759, $p < .01$. The differences in RTs mirror those for accuracy, with faster RTs on filler sentences than target sentences, t (41) = 8.031, $p < .01$, faster RTs on both correct, t (41) = 8.684, $p < .01$, and incorrect, t (41) = 3.783, $p < .01$ fillers as compared to target sentences, and finally, faster RTs on correct fillers than on incorrect fillers, t (41) = 7.378, $p < .01$. The mean accuracies and RTs for the target, incorrect, and correct fillers are displayed in Figures 34 and 35.

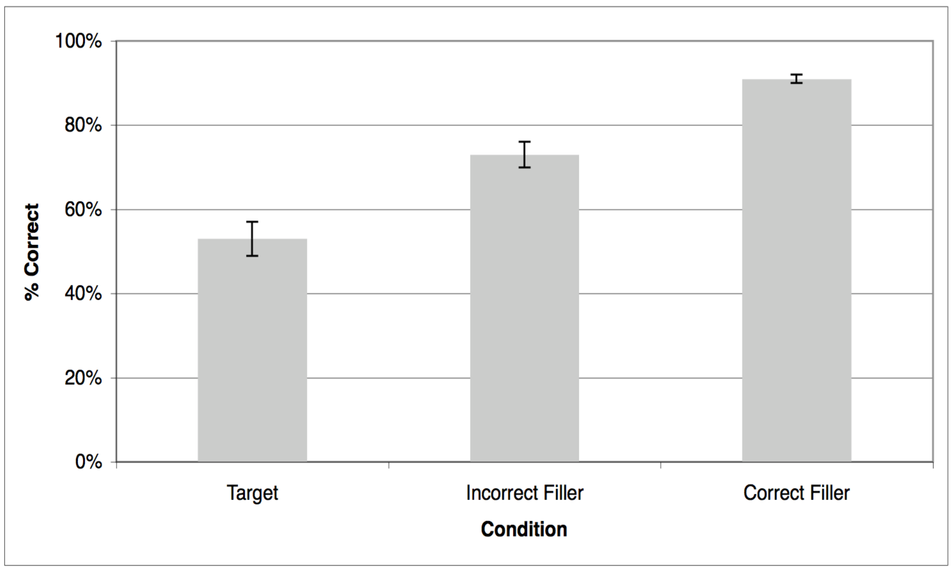


Figure 34. ENS accuracy (%) on target, incorrect filler, and correct filler sentences

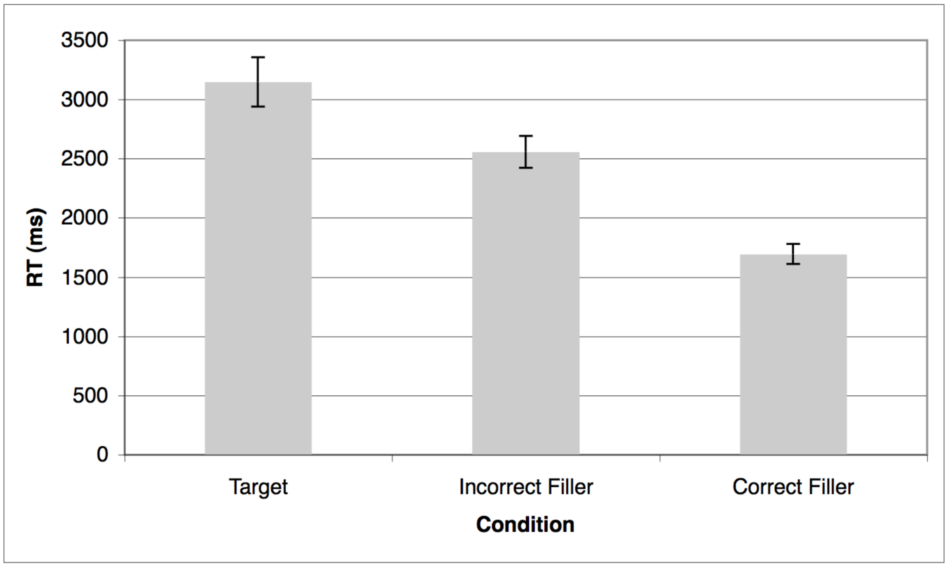


Figure 35. ENS RT (ms) on target, incorrect filler, and correct filler sentences

The subject and item analyses for accuracy showed significant main effects of both Cue, *F1* (1, 41) = 24.878, $p < .01$, *F* 2(1, 40) = 18.077, $p < .01$, and Gender *F1* (1, 41) = 5.615, $p < .05$, *F* 2(1, 40) = 5.461, $p < .05$. ENS participants achieved higher accuracy (*M* = 59%, *SD* = 27%) on sentences in which a gender cue was provided than on sentences in which no gender cue was provided (*M* = 46%, *SD* = 25%) and on sentences in which a feminine noun was modified by a masculine adjective (*M* = 57%, *SD* = 26%) than on sentences in which a masculine noun was modified by a feminine adjective (M = 49%, SD = 27%).

There were no other main effects or interactions that were significant in both the subject and item analyses; however, a three-way interaction between Distance, Cue, and Gender was significant in the subject analysis, *F*1 (1, 41) = 4.862, $p < .05$. Figure 36 shows that ENSs achieved higher accuracy on sentences in which a feminine noun was modified by a masculine adjective in both the close and far condition when no cue was provided, but only in the close, and not the far, condition when a cue was provided. There were no significant main effects or interactions for RT in both the subject and item analyses, although there was a main effect of Gender RT in the subject analysis, $F1$ (1, 41) = 10.2, $p < .01$, with faster RTs on sentences with a feminine noun modified by a masculine adjective (M = 2911, SD = 1160) as compared to sentences with a masculine noun modified by a feminine adjective (M = 3474, SD = 1908).

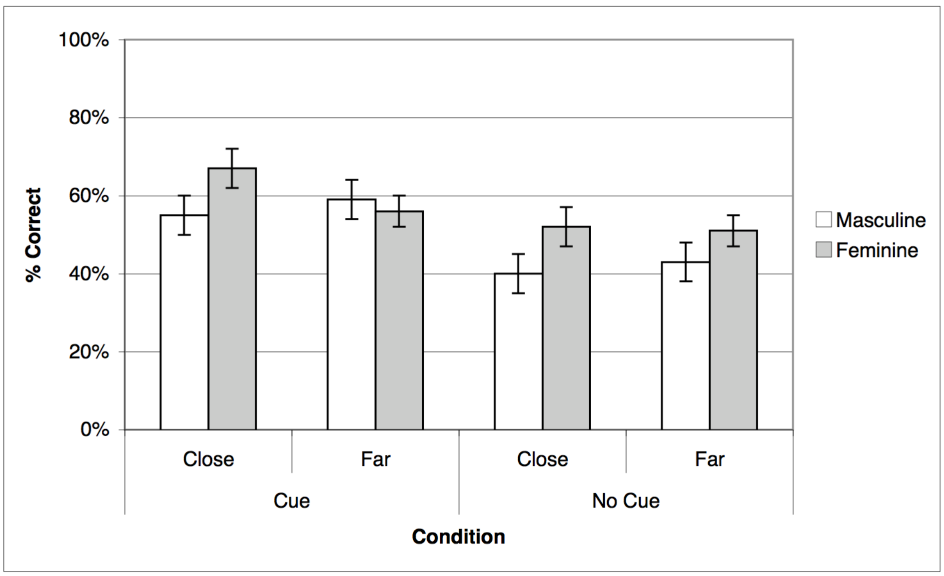


Figure 36. ENS accuracy (%) on target sentence conditions

Finally, when low and high span groups were included in the item analysis as the within-items variable, there were no significant main effects or interactions. Interestingly, whereas the other language groups showed higher overall accuracy for the high span groups, the ENS high span participants did not show a difference in accuracy between high (M = 53%, SD = 17%) and low (M = 58%, SD = 15%) span groups. However, this result may be explained by the fact that the low span group had known French for more years (31 years, range 10-55) and spent more time in France (17 years, range .75-41) compared to the high span group who had known French for fewer years (19 years, range 4-47) and spent less time in France (9 years, range .6-22).

An item analysis for filler sentence accuracy and RT, with span as the within-items variable, revealed no effect of span, even when the correct and incorrect fillers were considered independently.

Overall, the ENS participants performed similarly to the SNS and DNS participants, with lowest accuracy on the target sentences and highest accuracy on correct filler sentences. The next section will provide an overview comparison of performance by all four language groups, as well as a discussion of these findings.

8.5.3 Discussion

The grammaticality judgment task was designed to investigate NS and NNS gender agreement accuracy during online processing. The 48 target sentences contained noun-adjective gender agreement errors, with the inaccurate adjective occurring either directly after (close condition) or several words down (far condition) from the target noun. Furthermore, half the sentences contained a gender-marked determiner (cue condition) and half the sentences did not provide a gender cue for the target noun (no cue condition). The 92 filler sentences were either correct or contained an error of complex negation, *avoir* vs. *être* verb form, or subject-verb agreement. It was expected that the FNS participants would perform well on the target sentences, regardless of condition, whereas NNS performance would depend on L1-L2 gender-system similarity, such that the SNSs would achieve the most native-like accuracy and the ENSs the least. Furthermore, noun-adjective distance, the availability of gender cue, and WM capacity were expected to play a role for the DNS and ENS participants, but not for the SNSs or FNSs.

All participants achieved highest accuracy on the correct filler sentences. For the FNSs, mean accuracy scores were higher for the target sentences (*M* = 94%, *SD* = 7%) than for the incorrect filler sentences (*M* = 88%, *SD* = 6%), indicating that identifying gender agreement errors in this task was easier than identifying the other types of errors found in the filler sentences. However, for all three NNS groups, mean accuracy scores were lowest for the target sentences, with at-chance performance for the ENSs (*M* = 53%, *SD* = 25%) and SNSs (*M* = 50%, *SD* = 23%), and slightly above chance for the DNSs (*M* = 62%, *SD* = 22%). A one-way ANOVA showed that the four groups differed significantly in target sentence accuracy, F (3, 134) = 36.065, $p < .01$. Tukey post-hoc comparisons showed that accuracy for the FNS group was significantly better than accuracy for the SNS, DNS, and ENS groups ($p < .01$ for all comparisons), but the SNS, DNS, and ENS groups did not differ significantly from each other. Figure 37 displays the target, incorrect filler, and correct filler sentence accuracy (%) for all four language groups.

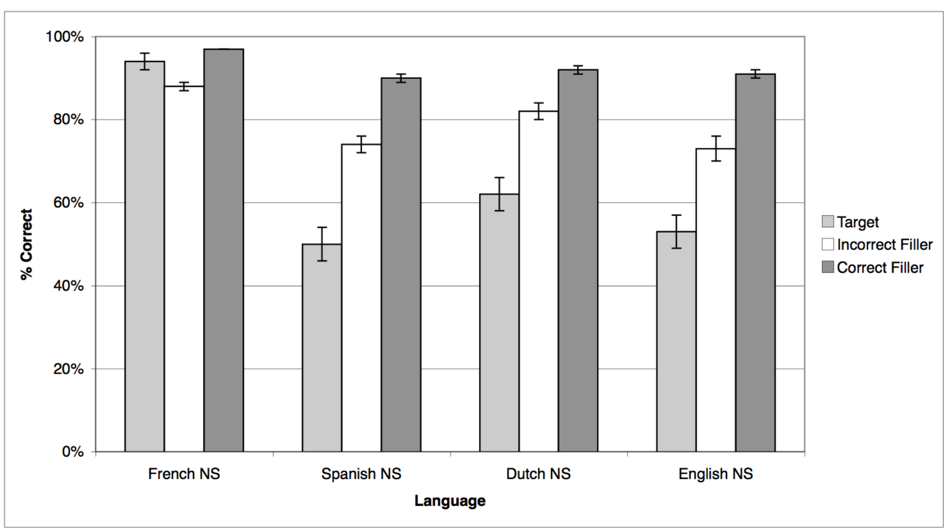


Figure 37. Accuracy (%) on target, incorrect filler, and correct filler sentences for all four language groups

To determine whether the differences between target, incorrect filler, and correct filler sentences were due to a yes-bias, a sensitivity index, d', was calculated. A series of paired samples t-tests indicated that, despite a potential yes-bias, sensitivity to the gender errors in the target sentences was significantly different than sensitivity to the errors in the incorrect filler sentences for all four language groups (FNSs: t (20) = 3.799, $p < .01$; SNSs: t (36) = -6.779, $p < .01$; DNSs: t (37) = -8.309, $p < .01$; ENSs: t (41) = -6.279, p < .01). Furthermore, a series of paired samples t-tests compared gender error sensitivity of the three NNS groups to that of the FNS group. With a Bonferroni correction for multiple comparisons, results indicated that, despite a potential yes-bias, all NNS groups demonstrated less sensitivity than the FNSs to the gender errors in the target sentences (FNS vs. SNS: t (56) = 12.545, $p < .0167$; FNS vs. DNS: t (57) = 8.589, $p < .0167$; FNS vs. ENS: t (61) = 10.084, $p < .0167$).

Although one possible explanation for the NNS low target sentence accuracy is that the NNSs do not know the gender of the target nouns, despite the cue provided in half of the target sentences, results from the gender assignment post-test indicate that this was not the case; the NNS participants achieved high gender assignment accuracy, demonstrating their gender knowledge. Figure 38 shows the gender assignment task accuracy (%) for each language group on the nouns that appeared in the target sentences. A significant difference among the four language groups was found in a one-way ANOVA, F (3, 134) = 11.044, $p < .01$. Although Tukey post-hoc comparisons showed that accuracy for the FNS group was significantly better than accuracy for the SNS, DNS, and ENS groups ($p < .01$ for all comparisons), and the NNS groups did not differ from each other, the NNSs clearly demonstrated knowledge of the target nouns' gender.[52]

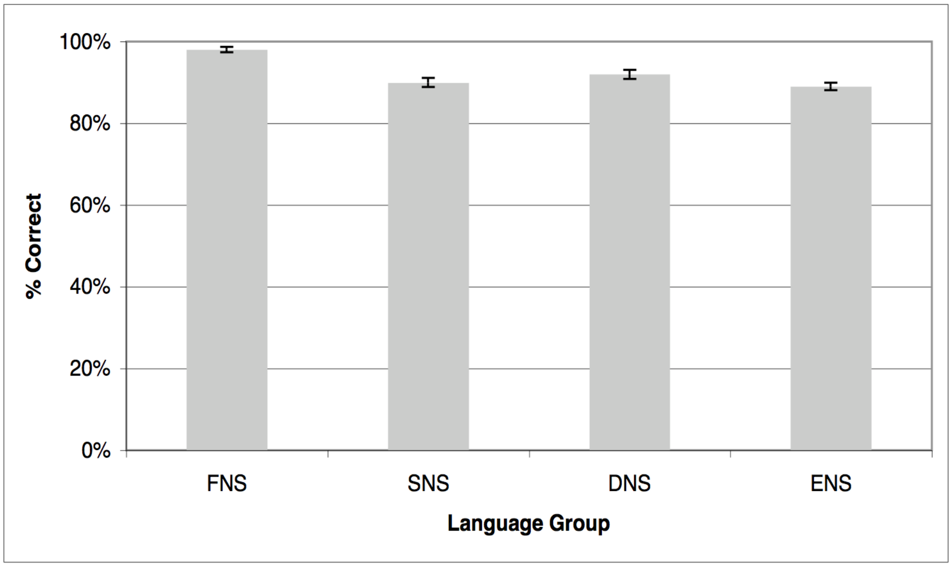


Figure 38. Accuracy (%) on grammaticality judgment target nouns in gender assignment post-test for all four language groups

The NNSs' high gender assignment accuracy confirms findings in previous studies that even highly proficient learners of French, despite accurate knowledge of a noun's gender, are rarely able to achieve native-like gender agreement ability. Furthermore, it appears that L1-L2 gender-system similarity does not facilitate L2 gender agreement, as the SNSs neither achieved native-like accuracy, nor achieved higher accuracy than the DNSs or ENSs.

Regarding RTs, the results indicate there was no speed-accuracy trade-off in this task, as there was a pattern for sentences with higher accuracy to have faster RTs and

[52] Target sentences for which participants did not correctly assign gender to the target noun in the gender assignment post-test were not excluded from the analyses. Because half the target sentences provided a gender cue, and because gender assignment accuracy was high (FNSs: M = 98%, SD = 3%, NNSs: M = 90%, SD = 7%), it is unlikely inaccurate gender knowledge is driving the low target sentence accuracy for the NNSs.

sentences with lower accuracy to have longer RTs. Because the RT analysis was secondary to the accuracy analyses, and the RT results corroborate the accuracy findings, the focus of this discussion section will be on accuracy results.

To facilitate discussion of target sentence accuracy performance across the four language groups, all significant main effects and interactions are presented in Table 33.

Table 33

ANOVA Results for Target Sentence Accuracy for All Four Language Groups

Variable	Language Group	Significance
Distance	DNS	Subject (*F*1)
Cue	SNS	Subject (*F*1)
	DNS	Subject (*F*1) & Item (*F*2)
	ENS	Subject (*F*1) & Item (*F*2)
Gender	FNS	Subject (*F*1) & Item (*F*2)
	SNS	Subject (*F*1) & Item (*F*2)
	DNS	Subject (*F*1) & Item (*F*2)
	ENS	Subject (*F*1) & Item (*F*2)
Distance x Gender	DNS	Subject (*F*1)
	ENS	Subject (*F*1)
Distance x Cue	---	---
Cue x Gender	---	---
Distance x Cue x Gender	DNS	Subject (*F*1) & Item (*F*2)
	ENS	Subject (*F*1)
Span	FNS	Item (*F*2) with Span
	SNS	Item (*F*2) with Span
	DNS	Item (*F*2) with Span

Based on the results presented in Table 33, it is clear that the distance between the target noun and modifying adjective neither facilitated nor inhibited gender agreement accuracy for any of the participants. The DNSs showed a main effect of Distance; however, it was only significant in the subject analysis and was subsumed by a three-way interaction with Cue and Gender that was significant in both the subject and item analyses.

Cue, however, appears to play a role in gender agreement accuracy, at least for the ENS participants, who achieved higher accuracy when a gender cue was provided than when no gender cue was provided. SNS and DNS participants also showed a main effect of Cue; however, for the SNSs, it was only significant in the subject analysis, and for the DNSs, Cue interacted with both Distance and Gender, with the presence of a gender cue improving accuracy only on sentences in the far condition when the target noun is masculine. However, the presence of a gender cue did result in faster RTs for the DNSs. Overall, it can be concluded that the presence of a gender cue clearly facilitates noun adjective gender agreement accuracy only for the ENSs; the pattern is less robust for the SNS and DNS participants.

A main effect of Gender was found for all four language groups in both the subject and item analyses. Although this main effect was complicated by a three-way interaction with Distance and Cue for the ENSs in the subject analysis and the DNSs in both the subject and item analyses, an interesting pattern emerged. For the FNSs, accuracy was higher on sentences in which a masculine noun is modified by a feminine adjective, whereas for all three NNS language groups, accuracy was higher on sentences in which a feminine noun is modified by a masculine adjective. Before drawing conclusions about this finding, it is necessary to consider gender assignment accuracy to determine whether this pattern is merely a reflection of the participants' differing knowledge of masculine and feminine nouns' gender. Figure 39 presents gender assignment post-test accuracy for the target sentence masculine and feminine nouns. Combining accuracy on both masculine and feminine nouns, all four language groups achieved high accuracy (FNSs: M = 98%, SD = 3%; SNSs: M = 90%, SD = 7%; DNSs: M = 92%, SD = 7%, ENSs: M = 89%, SD = 6%), and despite the slightly lower gender assignment scores for the NNSs, accuracy for both masculine and feminine nouns was high compared to the low accuracy scores on grammaticality judgment task target sentences.

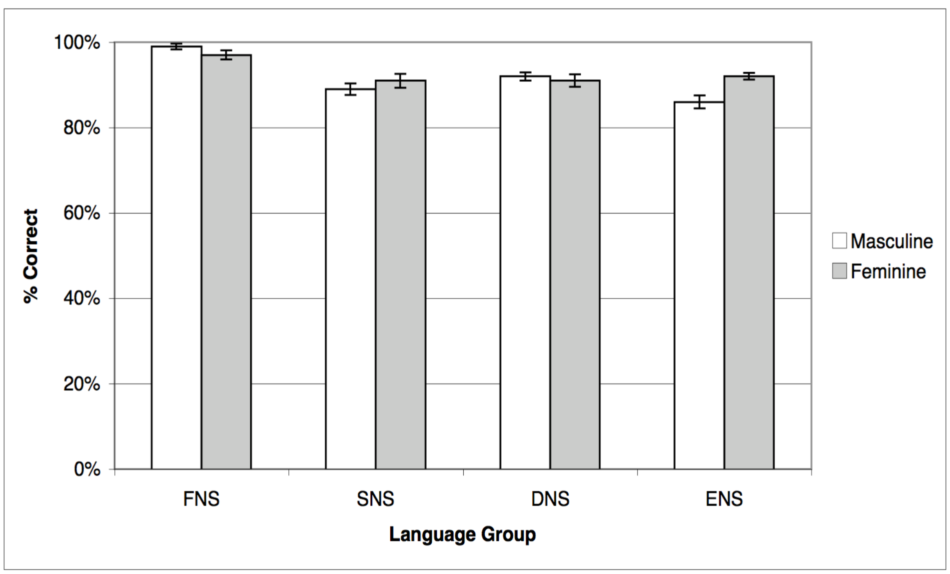


Figure 39. Accuracy (%) on masculine and feminine target nouns in gender assignment post-test for all four language groups

In a set of paired-samples t-tests, only the ENSs showed a significant difference between masculine and feminine noun gender assignment accuracy, t (41) = 4.078, p <

.01. Therefore, a discrepancy in ability to assign gender to masculine and feminine nouns does not explain the main effect of Gender, nor the opposite pattern in accuracy on masculine and feminine nouns for NS and NNS participants.

Another possible explanation for the opposing NS-NNS main effect of Gender is the participants' perception of the base, or default, form of the adjective. For NNSs, the masculine form may be considered the default, in which case the NNSs may assume they do not need to do anything to the modifying adjective. For NSs, however, the masculine form of the adjective is not necessarily the base form. Valdman (1976) specifically proposed that the masculine is not the base form, but rather there is an underlying form from which the masculine and feminine forms are derived, with the masculine form consisting of a "zero ending" that prevents the final phoneme of the base form from being realized. In this case, it is the feminine form, not the masculine, that is considered the base, or default form, and the incorrect use of the base (feminine) form may serve as a stronger trigger for ungrammaticality than the modified (masculine) form. Under this explanation, both NS and NNS accuracy is higher when the perceived base form of the adjective incorrectly modifies a noun, but the perception of which form is the base form differs. Regardless of the underlying explanation, for NSs, the presence of the adjective's final phoneme is a more obvious error than its absence, whereas for NNSs, the absence of the adjective's final phoneme is a more obvious error than its presence.

For the FNS, SNS, and DNS participants, a main effect of WM span emerged for overall accuracy on the grammaticality judgment task target sentences in the item analysis with Span as the within-items variable, with the high span groups achieving higher accuracy than the low span groups. The ENS low span group performed similarly to the high span group; however, this may be explained by the fact that the low span group had known French for longer and spent more time in France than the high span group. Overall, based on the high FNS and uniformly low NNS accuracy scores, Span (either high or low) does not provide an impressive advantage for any of the language groups.

In addition to the analyses carried out separately for each language group, one-way repeated-measures ANOVAs for accuracy ($F1$ and $F2$) were performed with the three NNS groups combined. The goal of these analyses was to determine whether one NNS language group had an advantage over another. For example, do NSs of Spanish have an advantage in French L2 gender agreement over NSs of Dutch or English? In the subject analysis ($F1$), with Distance, Cue, and Gender as the within-subjects variables, and Language as the between-subjects variables, significant interactions between Language and Distance, $F1\ (2, 114) = 3.267$, $p < .05$, and Language and Cue, $F1\ (2, 114) = 4.925$, $p < .01$, emerged. In the item analysis ($F2$), with language as the within-items variable, and Distance, Cue, and Gender as the between-items variables, there was a main effect of Language, $F2\ (2, 80) = 24.369$, $p < .01$, and a significant interaction between Language and Cue, $F2\ (2, 80) = 4.065$, $p < .05$. There was no significant interaction between Language and Distance, as found in the subject analysis. As discussed above, the DNSs achieved higher accuracy than the SNS and ENS participants, although this difference was not significant in a one-way ANOVA. Regarding the interaction between Language and Cue, the DNSs benefited more from a gender cue than the SNS or ENS participants, as shown in Figure 40; however, as discussed in Section 8.5.2.3, for DNSs, Cue was involved in a three-way interaction with Distance and Gender, making it difficult to conclude that DNSs benefited from a gender cue. Overall, none of the NNS language groups had a clear advantage over another.

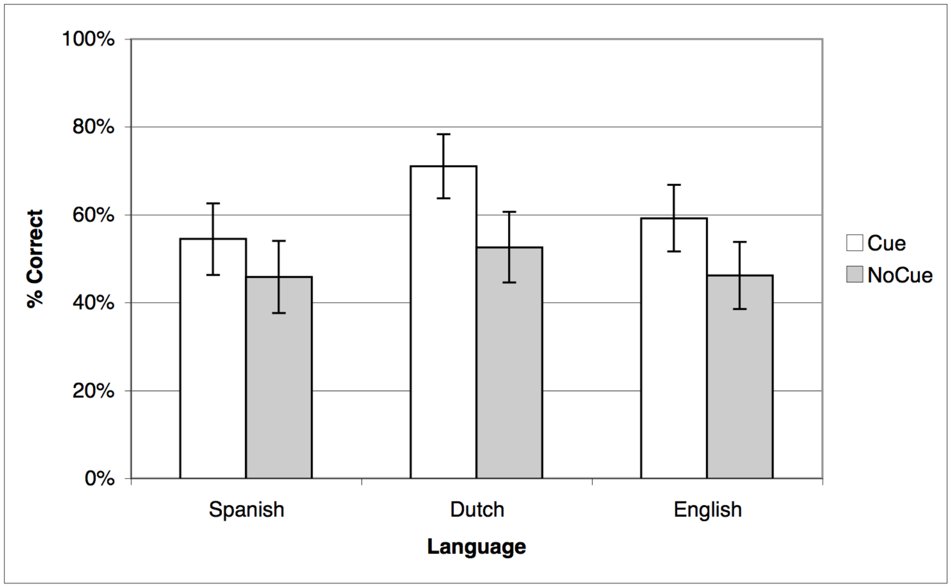


Figure 40. SNS, DNS, and ENS accuracy (%) on sentences in the Cue condition

Finally, accuracy on the target sentences was included in the correlation matrix for each language group. Although a main effect of Span was found in the item analysis, in the correlation matrix analysis with Span as a continuous variable there was no significant correlation with target sentence accuracy for any of the four language groups. Specifically, WM span scores did not correlate with overall target sentence accuracy, or accuracy in the Distance, Cue, or Gender conditions. The lack of significant correlations validates the conclusion that WM capacity does not provide an advantage on noun-adjective agreement accuracy for any of the four language groups. Furthermore, Span did not correlate with accuracy on the filler sentences.

Target sentence accuracy did, however, correlate with target noun gender assignment post-test accuracy for the three NNS language groups, SNS: $r(37) = .641$, $p < .01$; DNS:

$r(38) = .842$, $p < .01$; ENS: $r(42) = .584$, $p < .01$). That is, NNS participants who achieved higher accuracy on the gender assignment post-test (for grammaticality judgment target nouns) achieved higher accuracy on the target sentences. Accuracy on the target sentences also correlated with AO for the SNSs, $r(37) = -.370$, $p < .05$, and ENSs, $r(42) = -.376$, $p < .05$. SNS and ENS participants with an earlier AO achieved higher accuracy, suggesting that gender agreement is age sensitive. That the DNS group did not show a corresponding correlation is probably due to their limited range for AO. For the SNSs only, target sentence accuracy correlated with the number of years a participant has known French, $r(37) - .379$, $p < .05$; the longer a SNS has known French, the higher the target sentence accuracy. However, this correlation may be a function of the overall fewer number of years the SNS participants have known French. In other words, it may be that this correlation appeared due to the few participants who have only known French for a few years, as compared to the overall greater number of years known in the DNS and ENS groups. Neither age nor the number of years spent in France correlated with target sentence accuracy for any of the NNS groups. And finally, looking at the three target sentence conditions (Distance, Cue, Gender), none correlated with age, AO, the number of years a participant has known French, or the number of years spent in France. In sum, there is no clear pattern or any obvious predictor of what facilitates gender agreement accuracy for NNSs. Overall, all NNSs performed poorly on the target sentences, with no indication of participant characteristics that might facilitate performance.

8.5.4 Conclusion

The hypotheses specific to the grammaticality judgment task are repeated below for the reader's convenience.

2a. Spanish learners of French will achieve near-native accuracy on gender agreement.

2b. Dutch learners of French will not achieve near-native accuracy on gender agreement.

2c. English learners of French will not achieve near-native accuracy on gender agreement.

3a. The availability of external gender cues will not facilitate gender agreement accuracy for Spanish learners of French.

3b. The availability of external gender cues will facilitate gender agreement accuracy for Dutch learners of French.

3c. The availability of external gender cues will facilitate gender agreement accuracy the most for English learners of French.

4a. WM span will not be correlated with gender agreement accuracy for Spanish learners of French.

4b. WM span will be correlated with gender agreement accuracy for Dutch learners of French.

4c. WM span will be correlated with gender agreement accuracy for English learners of French.

Regarding hypotheses 2a-2c, none of the three NNS language groups achieved near-native gender agreement accuracy. Rather, their performance was approximately at chance. The availability of external gender cues did not facilitate gender agreement accuracy for SNSs, as predicted in hypothesis 3a; however, gender cues also did not facilitate gender agreement accuracy for DNSs, contrary to hypothesis 3b. Hypothesis 3c was supported: the availability of gender cues facilitated gender agreement accuracy for ENSs. Finally, high WM span unexpectedly facilitated overall gender agreement accuracy for the SNS participants (hypothesis 4a). Whereas high WM span was expected to, and did, facilitate overall gender agreement accuracy for DNSs (hypothesis 4b), it did not for the ENS participants (hypothesis 4c). Furthermore, that the FNSs also showed facilitation in the high span groups suggests that WM capacity plays a role in the overall ability to complete the task, rather than facilitating gender agreement accuracy specifically. Finally, despite a main effect of WM span in the item analysis, it did not correlate with accuracy in any of the target sentence conditions or with accuracy on the filler sentences in the correlation matrix.

Overall, none of the NNSs realized gender agreement similarly to FNSs, and gender agreement accuracy scores were surprisingly low for all three NNS language groups. Other than facilitation with the presence of a gender cue for the ENSs, no clear patterns of facilitation appeared for Distance and Cue conditions. The NNSs consistently achieved higher accuracy on sentences in which a feminine noun was modified by a masculine adjective, as compared to sentences in which a masculine noun was modified by a feminine adjective, a pattern that is opposite for the FNSs. Finally, WM capacity does not appear to play a role for NNSs in facilitating gender agreement accuracy.

Chapter 9: General Discussion and Conclusion

Mastering an L2 grammatical gender system is a difficult feat, even for highly advanced language learners. Although NNSs have little difficulty assigning gender to L2 nouns, realizing appropriate gender markings throughout a sentence remains a challenge. Recent research on gender representation suggests NSs store gender as an inherent property of a noun, allowing for accurate gender agreement; however, NNSs' inability to achieve native-like gender agreement suggests they do not store L2 grammatical gender in a native-like fashion. Although there is evidence that L1-L2 gender-system similarity facilitates L2 gender processing (Sabourin & Stowe, 2008), no studies to date have examined the influence of the L1 gender system on both gender representation and gender processing.

The dissertation research set out to investigate NNS gender representation and gender processing, with the specific aim of examining the role of L1. In addition, because it was predicted that NNSs whose L1 does not have a gender system similar to French would rely on external factors during gender processing, the roles of gender cues and non-linguistic processing constraints were also considered.

9.1 Summary of Findings

9.1.1 Gender Priming Task

The goal of the gender priming task was to determine whether participants store L2 grammatical gender as an inherent property of a noun, as demonstrated by gender priming effects. Participants were presented with determiner primes that were congruent, incongruent, or neutral with respect to the gender of the name of a target picture. The following general hypothesis was proposed for this task:

1. Spanish and Dutch, but not English, learners of French will represent grammatical gender similarly to French NSs.

The FNS controls exhibited priming effects in that they were faster to name target pictures when the prime-target pair was congruent than when the pair was incongruent. This finding replicates previous research in which NSs show congruency effects in gender priming tasks (Alario et al., 2004) and gender distraction tasks (Schriefers, 1993), indicating that a gender-marked prime activates a gender node, which, in turn, facilitates (or interferes with) production of a noun of the same (or different) gender. The interference effects found by the FNSs in this task are taken as evidence of nodal links between a gender node and a lemma, the key component of the NS gender storage and nodal system proposed in Levelt's model of speech production (Levelt, 1989; Levelt, Roelofs, & Meyer, 1999).

However, none of the three NNS language groups showed clear gender priming effects, even when item characteristics (noun ending ambiguity and word frequency) and participant characteristics (AO, number of years spent in France, number of years participant had known French, and WM) were taken into account. The first hypothesis, therefore, was not supported. Regardless of advanced L2 proficiency and L1-L2 gender-system similarity, none of the NNS language groups showed evidence of a native-like gender storage and nodal system.

9.1.2 Grammaticality Judgment Task

The purpose of the grammaticality judgment task was to examine NNSs' ability to realize L2 gender agreement during online processing. Using an RSVP paradigm, sentences were presented one word at a time, with target sentences containing noun-adjective gender agreement errors. The adjectives occurred either directly after the target noun, or several words down, and either with or without a gender-marked determiner before the target noun. The participants indicated detection of an error by button push at any point during presentation of the sentence. The following general hypotheses were proposed for this task:

2. Spanish, but not Dutch or English, learners of French will realize gender agreement similarly to French NSs.
3. Dutch and English, but not Spanish, learners of French will rely on external gender cues during gender processing.
4. For Dutch and English, but not Spanish, learners of French, WM span will be correlated with accurate gender agreement.

The FNSs achieved high accuracy on both the target and filler sentences, demonstrating that identifying noun-adjective agreement errors did not pose a problem. Furthermore, FNSs performed equally well regardless of noun-adjective distance or the presence/absence of a gender cue. However, the FNSs did show an effect of Gender, with higher accuracy on sentences in which a masculine noun was incorrectly modified by a feminine adjective as compared to sentences in which a feminine noun was incorrectly modified by a masculine adjective. Although accuracy was high in both conditions, the difference suggests that the presence of the final phoneme on an adjective is a more salient error than the absence of the final phoneme on the adjective.

The NNSs, however, all performed poorly on the target sentences, with the SNS and ENS participants performing approximately at chance (50% and 53%, respectively), and the DNSs only slightly better (62%). Furthermore, whereas the FNSs performed better on the target sentences than on the incorrect filler sentences, this was not true for the NNSs. The target sentences were more difficult for the NNSs than the incorrect filler sentences, indicating that gender agreement errors pose a distinct challenge for NNSs. Regardless of L1-L2 gender-system similarity, none of the NNS participants realized gender agreement similarly to FNSs. Furthermore, there was no distinction between the NNS language groups based on L1-L2 gender-system similarity; the SNSs, who have a congruent L1-L2 gender system, did not perform better than the ENSs, who have a minimal L1 gender system. The second hypothesis, therefore, was only partially supported, with none of the three NNS language groups realizing gender agreement similarly to French NSs.

The presence of a gender cue was most beneficial for the ENSs, who achieved higher accuracy when a cue was provided. SNSs also achieved higher accuracy when a cue was provided, but this finding was not robust, as it was only significant in the subject analysis. The DNSs also showed some benefit from the presence of a gender cue on accuracy, but the main advantage was apparent in RTs, with faster responses on sentences in which a cue was provided. Overall, therefore, the third hypothesis was partially supported, with only the ENSs showing a clear reliance on external gender cues during processing. Although SNS and DNS participants did benefit from the presence of gender cues, the pattern was neither consistent nor robust.

Finally, the fourth hypothesis was also only partially supported. As predicted, DNSs showed an effect of WM span: high span participants achieved higher accuracy on the target sentences than low span participants. However, the same was true for the SNSs and FNSs, whereas the ENSs showed no effect of span, with low span participants performing similarly to high span participants. It is important to keep in mind, though, that WM span did not correlate with accuracy on target sentences or any of the target sentence conditions in the correlation matrix, indicating that WM capacity did not play a significant role in gender agreement for any of the participants.

9.2 Implications

Overall, notwithstanding weak evidence of gender priming effects, the NNSs in this study neither represent L2 grammatical gender nor realize gender agreement similarly to NSs, regardless of their L1-L2 gender-system similarity. Even if the inconsistent gender priming effects are to be taken as evidence of native-like gender representation, this did not provide an advantage during gender processing given the at-chance performance by both SNS and ENS participants, and the only slightly higher performance by the DNSs, on grammaticality judgment task target sentences. Furthermore, external cues and non-linguistic processing constraints did not uniformly benefit learners whose L1 does not have a gender system similar to that of French. Despite advanced proficiency, the NNSs did not show evidence of native-like gender representation, and they performed exceptionally poorly on noun-adjective gender agreement on the online processing task.

These results are consistent with previous findings, as L2 gender agreement is notoriously difficult and no studies to date have shown evidence of native-like L2 gender representation (Bordag et al., 2006; Guillelmon & Grosjean, 2001). That the NNSs did not show evidence of native-like gender representation indicates that they have not developed the gender-nodal system that allows for automatic activation of gender, as proposed by Levelt's model of speech production (Levelt, 1989; Levelt, Roelofs, & Meyer, 1999). Although the NNSs' high gender assignment accuracy demonstrates that they have created a store of French gender knowledge, this store may be external from grammatical information at the lemma level. In other words, NNS gender knowledge does not benefit from automatic activation, thus resulting in non-native-like L2 gender processing.

If, indeed, only NSs are able to develop a gender-nodal system, the question as to the role of age effects in acquiring grammatical gender arises. Given that even the DNSs, whose AO was the earliest of the three NNS groups, did not show evidence of gender priming, it is likely that gender acquisition is age sensitive and that the cutoff for achieving a native-like gender system is quite young. Data showing that even early immersion students (AO of 5-6 years) are not native-like in their L2 gender system (Harley, 1979; Lapkin & Swain, 1977) support the notion of an early cutoff. If FNSs master the gender system around the age of three (Karmiloff-Smith, 1979), and NNSs with an AO of 5 years are not able to become native-like, it is reasonable to infer that the critical period for acquiring a native-like gender system ends around the age of 3 or 4.

Although there is abundant evidence that even advanced NNSs do not achieve native-like gender agreement (Bartning & Schlyter, 2004; Dewaele & Véronique, 2000, 2001), Sabourin and Stowe (2008) found evidence for L1-L2 gender-system similarity facilitating gender agreement on a grammaticality judgment task, a finding that was not confirmed in the current study. The surprisingly low gender agreement performance by the SNSs raises the question as to why the SNSs did not show an advantage over the ENSs considering the similarity between the Spanish and French gender systems.

One possible explanation is that the similarity between the French and Spanish gender systems interfered with, rather than facilitated, SNS performance. If a target noun in the grammaticality judgment task is feminine in French, but masculine in Spanish, a SNS participant may apply masculine gender agreement throughout the sentence based on the L1 gender representation of the target noun, thus resulting in low accuracy. Of the 48 grammaticality judgment task sentences, 15 (30%) had a French-Spanish gender mismatch. However, when these sentences were removed, SNS accuracy on the target sentences only improved to 56% (from 50% accuracy when the mismatch sentences were included). The SNSs performed poorly, only slightly above chance, even when the target nouns were the same gender in both the L1 and L2. Thus, the L1-L2 gender mismatch does not account for the low SNS accuracy on this task.

A second possible explanation lies in the correspondence of the French and Spanish gender systems. Although the gender systems appear similar, two distinct differences may prevent an advantage for SNS learners of French. The first difference is the rule system for assigning gender to nouns. The French system is governed by noun-ending rules, but the rules are opaque and somewhat unreliable. The Spanish system is also governed by noun-ending rules, but the rules are transparent and more reliable than those of French. A transparent, reliable L1 rule system may not be beneficial for making use of an opaque, unreliable L2 system. A second difference is the formation of adjectives. In Spanish, masculine adjectives are typically marked by an [o] ending and feminine adjectives by an [a] ending. Although this generalization is not without exception, as masculine adjectives may also end in a consonant and not all feminine adjectives end in [a], it often results in matching phonological noun and adjective endings, that is, a phonological pattern between the noun and adjective endings (examples 23 and 24).

(23) La cas<u>a</u> es blanc<u>a</u>

The house is white

(24) El libr<u>o</u> es blanc<u>o</u>

The book is white

In other words, the rules for assigning gender to nouns, the formation of adjectives, and noun-adjective gender agreement in Spanish are more transparent than in French. Again, the transparent processes used for Spanish gender agreement may not provide any advantage for making use of the opaque processes of French gender agreement.

This explanation is consistent with De Bot's (1992) claim that the processes used for carrying out gender agreement in the L1 are different from those used in the L2; however, it is inconsistent with Pienemann's (1998a, 1998b) claim that transfer of some L1 procedures can occur. Within De Bot's and Pienemann's theories, two scenarios regarding the current study are possible: (1) the Spanish and French gender systems are not similar enough for L1 processing procedures to transfer to the L2, putting the SNSs at the same disadvantage as ENSs in terms of mastering the French L2 gender system, or, (2) Spanish L1 procedures do transfer, but are not beneficial due to the gender system differences. Given that Sabourin and Stowe (2008) found an advantage for L1-L2 similarity in determiner-noun gender agreement, it may be that SNSs are able to transfer some of their L1 processing procedures, but this is not beneficial in noun-adjective gender agreement. Because adjectives and nouns do not always occur together, and, as opposed to determiners, adjectives carry semantic meaning, which may interfere with the processing of gender agreement, noun-adjective agreement may be more difficult than determiner-noun agreement. That is, this dissertation examined the most difficult aspect of gender agreement, and it may be that transfer does provide some advantage, for example, for determiner-noun agreement, as seen in Sabourin and Stowe.

In sum, the FNSs, but not the NNSs, showed evidence of a French gender storage-nodal system, as proposed in Levelt's model of speech production. In addition, NNSs, despite accurate gender knowledge, did not achieve native-like gender agreement. This finding suggests that L2 gender information is not available for L2 processing procedures in a way that allows for native-like processing; specifically, NNSs do not benefit from automatic activation of gender information during gender processing.

9.3 Additional Considerations

Before turning to directions for future research, it is important to address the limitations of the experimental tasks that may have influenced the results. First, as discussed in Chapter 8, there appears to be some characteristic(s) of the lists that results in list effects for the NNSs. A simple picture naming task comparing RTs across the three lists for both NS and NNSs would reveal whether the list effect occurs even when priming is not involved, and if it does, whether there are certain items driving the effect.

A second potential confounding factor in the gender priming task is L1 gender interference for the SNS and DNS participants. It may be that the presentation of a target picture simultaneously activates the L1 and L2 lemma, thus, activating the L1 gender information. Although the majority of target nouns were L1-L2 gender congruent for the SNSs (33 of 48), and would, therefore, be expected to facilitate gender priming, it may be that suppression of L1 gender information interacts with the time course of gender priming. For the DNSs, L1 gender is neither congruent nor incongruent with L2 gender; however, automatic activation of L1 gender may still interfere with the design of the priming task. Despite the efforts made to minimize L1 interference in the current study (i.e., including only highly proficiency L2 learners who were currently immersed in French), developing other gender node activation paradigms may lead to a more valid experimental task that can be used across L1 language groups.

Along these same lines, all NNS groups were highly multilingual, with many participants having studied languages, other than French, with gender systems. Participants' L3 gender systems may have interfered with performance on the gender priming task; it is, therefore, conceivable that all three NNS groups have developed a French L2 gender-nodal system, but L3 gender interference masked priming effects. But considering the difficulty NNSs have with French gender agreement, this scenario is unlikely.

In addition, although the FNSs showed gender priming effects, the task did not replicate exactly the findings of Alario et al. (2004), and further investigation into the role of the prime is necessary, for example, to determine what types of words serve as effective gender primes. In both Alario et al. and the current study, possessive pronoun primes (*mon*/*ma*) were not as effective as definite (*le*/*la*) and indefinite determiner (*un*/*une*) primes; however, it is unclear whether this is a function of prime frequency, with possessive pronouns being less frequent than definite and indefinite determiners, or whether it is a function of possessive pronouns carrying more meaning, and, thus, requiring more processing resources that potentially interfere with priming effects. To investigate what types of words make effective primes, one could develop a gender priming task that includes prime words with varying degrees of meaning and frequency, such as definite and indefinite determiners, possessive pronouns, subject pronouns, adjectives, and nouns.

Another consideration is the role of phonological noun endings in the gender priming task. Levelt's (1989) model indicates that phonological form is independent of gender processing; however, the FNSs in the current task showed a sensitivity to noun ending ambiguity, suggesting phonology does play a role during gender activation. It was not possible to determine whether this finding is a function of an unexpected component of the task, or representative of NS gender processing. Including noun ending ambiguity in a replication of Alario et al.'s (2004) priming task, or Schriefer's (1993) gender distraction task (although this would be difficult with Dutch noun stimuli) might shed light on this matter.

Turning to the grammaticality judgment task, the surprisingly low NNS accuracy raises some questions about the nature of the task. It is possible the RSVP paradigm is too difficult for NNSs and does not accurately represent their gender agreement ability. However, given that the FNSs did not have difficulty identifying gender agreement errors, and furthermore, achieved lower accuracy on the incorrect filler sentences than the target sentences, whereas the NNSs performed significantly better on the incorrect filler sentences, suggests that it is gender agreement, not the task itself, that is difficult for NNSs. It would be interesting to alert the NNSs to the types of errors to expect to determine whether explicitly paying attention to gender improves accuracy.

Finally, the low math accuracy in the O-Span task is a concern. As discussed in Section 8.3, a possible explanation is that the participants completed this task in their L1 while immersed in the L2. It is conceivable that, for some participants, this challenge affected their performance and that the span scores are not representative of their WM capacity. Although it was not possible to avoid L2 immersion, as it was an important criterion for participant inclusion, or to conduct the O-Span task in the L2, an O-Span task that required solving mathematical equations and recalling sets of letters, as opposed to words, may have minimized this limitation.

9.4 Conclusions and Future Directions

The current study adds to the growing body of literature suggesting that mastering an L2 gender system is exceptionally difficult, even for highly proficient L2 learners. This difficulty is likely rooted in NNSs' inability to store L2 grammatical gender as an inherent property of a noun, making automatic activation of gender information impossible. In addition to investigating the relationship between gender representation and gender processing, this study provides evidence that L1-L2 gender-system similarity does not facilitate gender agreement, nor do external factors, such as gender cues and WM capacity.

Future research should investigate the relationship between AO and L2 gender representation. An important question to ask is, at what age does an L2 learner lose the ability to develop a gender-nodal system similar to that of a NS, and is this cutoff consistent across L1s? Administering a gender priming task to monolingual NSs, and to early and late bilinguals (across a range of L1s), would address the role of the afe effects in developing a native-like L2 gender system.

It would also be informative to investigate more closely the role of L1-L2 gender system transfer. Although, in the current study, transfer did not seem to occur for any of the groups of participants, it may be that certain aspects of an L1 gender system do transfer. For example, a similar L1-L2 gender system may facilitate determiner-noun agreement, but not noun-adjective agreement. In addition, certain measures of gender agreement awareness may be more sensitive to L1 transfer than others. The RSVP paradigm in the current study revealed low gender agreement accuracy, but adding a physiological measure, such as ERPs, may indicate a sensitivity to gender agreement errors that is not apparent in a simple behavioral task. Previous ERP studies indicate L1 facilitation in gender agreement; however, future ERP studies should include the role of noun-adjective distance, the presence of external gender cues, and WM capacity, to determine how these variables interact to facilitate or interfere with L2 gender agreement.

Finally, the results of L2 gender research should be considered in a pedagogical context. If advanced L2 learners of French are unable to develop a native-like gender system in the L2, then it is unrealistic to expect lower–level learners of French to master this aspect of the language. That is, setting realistic expectations for students is an important component in developing effective and successful foreign language programs and language proficiency measures. Understanding the difficulty learners face in acquiring the French gender system will also allow teachers to help students develop effective strategies to overcome non-native-like gender agreement.

Appendix A

Pre-screening sentences

Type of error	Sentence	Discriminability Score
Contain errors in complex negation	*Heureusement, l'orage de la semaine dernière n'a pas fait aucun dégât matériel.	0.00
	*Mon camarade ne peut dire rien à sa mère quand elle se fâche.	.33
Contain errors in verb form (avoir vs. être)	*Ma grand-mère m'est offert une armoire l'année dernière.	0.00
	*Elles ont sorties du musée après y avoir passé deux heures.	.17
Contain errors in subject-verb agreement	*Leurs visages a révélé leur joie quand ils ont découvert le trésor ancien.	0.00
	*Au lieu d'y aller en voiture, les joueurs prend le train à leurs matchs de foot.	.33
	*Grâce à la gentillesse de l'infirmière, les malades reprend de courage.	0.00
Contain no errors	La maladie dont elle souffre terriblement n'a pas de remède.	.33
	À cause de la proximité des usines, la mer est malheureusement très sale.	.17
	Quand j'étais enfant, on se mettait généralement à table à dix heures.	0.00
	La petite fille adorait sa poupée, donc quand elle l'a perdue, elle était vraiment triste.	1.00
	Le haut fonctionnaire a patiemment interviewé les candidats pour le nouveau poste.	.33
	La nouvelle a été rapidement répandue dans toute la ville.	.33
	L'article présente brièvement les idées des deux professeurs qui ne s'entendent pas.	.33

Appendix B

Gender priming: target noun phoneme ambiguity

Phoneme	Surridge*	Lyster	Combined	Frequency (# of occurrences out of 9961 words)
o	97% Masc.	93% Masc.	Unambiguous (95% Masc.)	312
ɛ	90% Masc.	93% Masc.	Unambiguous (91.5% Masc.)	239
ʒ	94% Masc.	87% Masc.	Unambiguous (90.5% Masc.)	303
m	92% Masc.	82% Masc.	Unambiguous (87% Masc.)	249
R	75% Masc.	63% Masc.	Unambiguous (69% Masc.)	1507
z	90% Fem.	97% Fem.	Unambiguous (93.5% Fem.)	239
i	83% Fem.	68% Fem.	Unambiguous (75.5% Fem.)	523
ɔ̃	70% Fem.	71% Fem.	Unambiguous (70.5% Fem.)	1061
ʃ	66% Fem.	90% Fem.	Unambiguous (78% Fem.)	105
s	62% Fem.	79 % Fem.	Unambiguous (70.5% Fem.)	598
t	ambiguous	79% Fem.	Unambiguous (64.5% Fem.)	679
j	68% Fem.	65% Fem.	Ambiguous (66.5% Fem.)	143
p	ambiguous	64% Fem.	Ambiguous (57% Fem.)	66
l	ambiguous	54% Fem.	Ambiguous (52% Fem.)	561
e	ambiguous	53% Masc.	Ambiguous (51.5%)	1001
g	73% Masc.	61% Fem.	Ambiguous (conflicted gender)	54

* Surridge does not provide percentages for the phonemes deemed ambiguous; therefore, approximately 50% is assumed.

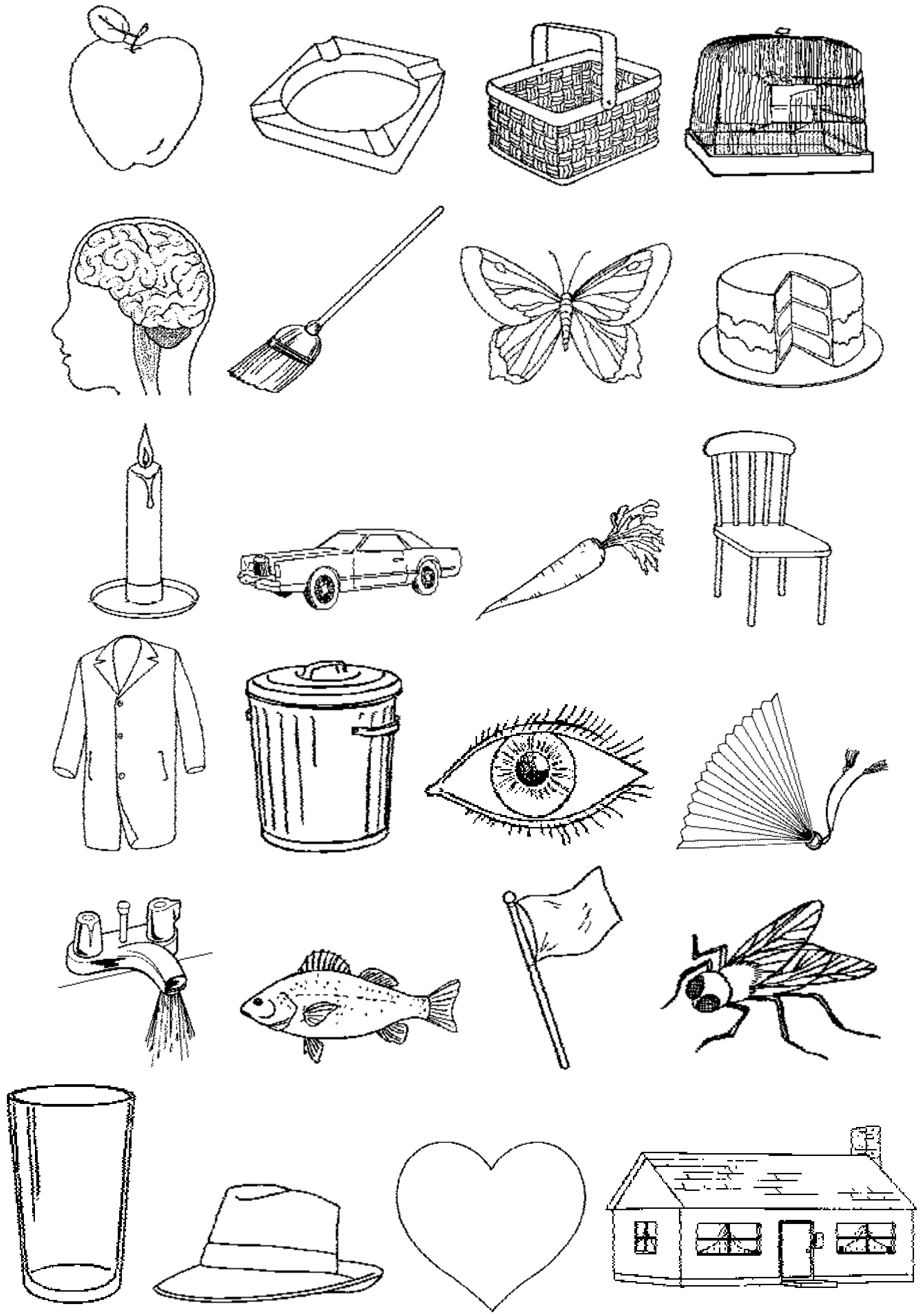
Appendix C

Gender priming materials (word list and pictures)

French (Intended Name)	English Translation	Gender	Name Agreement	Lexique Frequency (per million)	Final Phoneme	Phonological Ambiguity	Prime
			List 1 (1-16)				
ampoule	light bulb	F	100	9.81	l	Ambig	il/elle
balançoire	swing	F	100	2.52	R	Exception	mon/ma
cendrier	ashtray	M	93	8.59	e	Ambig	mon/ma
champignon	mushroom	M	100	7.89	ɔ̃	Exception	le/la
citrouille	pumpkin	F	93	2.81	j	Ambig	mon/ma
coeur	heart	M	100	300.21	R	Unambig	il/elle
crayon	pencil	M	100	18.64	ɔ̃	Exception	un/une
échelle	ladder	F	100	24.20	l	Ambig	un/une
écureuil	squirrel	M	100	7.73	j	Ambig	un/une
jupe	skirt	F	100	25.28	p	Ambig	le/la
maison	house	F	100	479.85	ɔ̃	Unambig	le/la
panier	basket	M	100	20.86	e	Ambig	il/elle
papillon	butterfly	M	100	13.32	ɔ̃	Exception	le/la
râteau	rake	M	100	1.44	o	Unambig	mon/ma
scie	saw	F	100	6.08	i	Unambig	un/une
tondeuse	lawn mower	F	100	1.75	z	Unambig	il/elle
			Mean = 99.1	**Mean = 58.19**			

French (Intended Name)	English Translation	Gender	Name Agreement	Lexique Frequency (per million)	Final Phoneme	Phonological Ambiguity	Prime
				List 2 (17-32)			
balai	broom	M	100	13.35	ɛ	Unambig	mon/ma
bougie	candle	F	100	16.96	i	Unambig	un/une
carotte	carrot	F	100	5.64	t	Unambig	mon/ma
chausette	sock	F	100	11.21	t	Unambig	il/elle
chaussure	shoe	F	100	32.66	R	Exception	le/la
cuillère	spoon	F	100	8.35	R	Exception	un/une
drapeau	flag	M	100	23.76	o	Unambig	le/la
mouche	fly	F	100	25.32	ʃ	Unambig	le/la
oeil	eye	M	100	504.15	j	Ambig	un/une
orteil	toe	M	100	6.69	j	Ambig	un/une
pomme	apple	F	100	42.37	m	Exception	il/elle
pouce	thumb	M	96	22.99	s	Exception	il/elle
robinet	faucet	M	100	9.83	ɛ	Unambig	le/la
selle	saddle	F	100	13.43	l	Ambig	mon/ma
tournevis	screwdriver	M	100	2.86	s	Exception	mon/ma
verre	glass	M	100	183.61	R	Unambig	il/elle
			Mean = 99.8	**Mean = 57.70**			

French (Intended Name)	English Translation	Gender	Name Agreement	Lexique Frequency (per million)	Final Phoneme	Phonological Ambiguity	Prime
			List 3 (33-48)				
bague	ring	F	100	24.01	g	Ambig	mon/ma
cage	cage	F	100	30.56	ʒ	Exception	il/elle
cerveau	brain	M	96	61.12	o	Unambig	mon/ma
chaise	chair	F	100	67.19	z	Unambig	le/la
chapeau	hat	M	100	65.69	o	Unambig	il/elle
clef	key	F	100	32.75	e	Ambig	un/une
couteau	knife	M	100	52.27	o	Unambig	le/la
cravate	tie	F	100	22.73	t	Unambig	le/la
étoile	star	F	100	47.41	l	Ambig	il/elle
éventail	fan	M	96	7.56	j	Ambig	il/elle
gateau	cake	M	100	36.82	o	Unambig	mon/ma
manteau	coat	M	93	45.77	o	Unambig	le/la
poisson	fish	M	100	50.49	ɔ̃	Exception	un/une
poubelle	garbage can	F	100	16.68	l	Ambig	mon/ma
sifflet	whistle	M	100	9.16	e	Ambig	un/une
voiture	car	F	96	306.54	R	Exception	un/une
			Mean = 98.8	**Mean = 54.80**			



Appendix D

Grammaticality judgment task: target noun phoneme ambiguity

Phoneme	Surridge*	Lyster	Combined	Frequency (# of occurrences out of 9961 words)
ã	99% Masc.	99% Masc.	Unambiguous (99% Masc.)	675
ʒ	94% Masc.	87% Masc.	Unambiguous (90.5% Masc.)	303
a	83% Masc.	85% Masc.	Unambiguous (84% Masc.)	259
R	75% Masc.	63% Masc.	Unambiguous (69% Masc.)	1507
z	90% Fem.	97% Fem.	Unambiguous (93.5% Fem.)	239
i	83% Fem.	68% Fem.	Unambiguous (75.5% Fem.)	523
ɔ̃	70% Fem.	71% Fem.	Unambiguous (70.5% Fem.)	1061
n	69% Fem.	82% Fem.	Unambiguous (75.5% Fem.)	348
v	69% Fem.	78% Fem.	Unambiguous (73.5% Fem.)	68
t	ambiguous	79% Fem.	Unambiguous (64.5% Fem.)	679
p	ambiguous	64% Fem.	Ambiguous (57% Fem.)	66
l	ambiguous	54% Fem.	Ambiguous (52% Fem.)	561
e	ambiguous	53% Masc.	Ambiguous (51.5% Masc.)	1001

* Surridge does not provide percentages for the phonemes deemed ambiguous; therefore, approximately 50% is assumed.

Appendix E

Grammaticality judgment sentences

Key for Target Sentences

Cue: Target noun is preceded by a gender-marked determiner
No Cue: Target noun is not preceded by a gender-marked determiner

Close: The adjective occurs directly following the target noun
Far: The adjective occurs at least four words after the target noun

Masc: The target noun is masculine (but the adjective is feminine)
Fem: The target noun is feminine (but the adjective is masculine)

Practice Sentences
Ce livre se vend à la librairie en ville, mais personne ne l'achète.
Le couple s'installe confortablement dans l'avion pour le long trajet à travers l'Atlantique.
Le jeune homme a hésité avant de présenter sa fiancée à sa famille.
*Ce matin mon frère me brusquement a réveillé car il voulait partir tôt.
*Si on travaille bien, on faire une bonne récolte cette année.
*L'institutrice a remercié les élèves qui lui sont envoyé des fleurs.

Cue + Close + Masc.
*Le prince a un nez laide, comme tous les hommes de sa famille.
*L'acteur continue à recevoir le courrier suspecte, même après l'arrestation du criminel.
*Le sable brulante ne gêne pas les lézards qui habitent dans le désert.
*La commode a un tiroir profonde; c'est là où elle met tous ses bijoux.
*L'appartement a un toit neuve, donc le propriétaire ne s'inquiète plus quand il pleut.
*Sa mère a préparé un repas délicieuse, même après une longue journée de travail.

Cue + Close + Fem.
*La dame ne voulait que de la toile élégant pour faire recouvrir ses canapés.
*L'enfant malin a laissé tomber une assiette précieux avec un éclat de rire.
*La boîte lourd qui se trouve dans le grenier appartient à ma mère.
*Tout le monde doit faire face à au moins une épreuve important dans la vie.
*L'étudiant a peur de faire une erreur évident dans son cours de français.
*Une église silencieux attire des gens qui cherchent un endroit où réfléchir tranquillement.

No cue + Close + Masc.

*Le pilote préfère les vols courtes car il s'ennuie facilement.

*La dame a choisi des meubles décoratives pour l'entrée du château.

*Le prêtre entend les gens qui racontent leurs péchés insignifiantes tous les jours.

*Elle met des croissants froides sur le radiateur pour les réchauffer.

*Le jeune étudiant n'a pas fait de cauchemar affolante depuis son enfance.

*Le ciel est plein de nuages blanches qui empêchent le soleil de percer.

No cue + Close + Fem

*Les enfants adorent leur école charmant et ils y vont joyeusement.

*Les équipes irlandais ont gagné tous les matchs de foot cette année.

*L'épée brillant se trouve actuellement au musée à Rome.

*Le patient s'imagine qu'il a des maladies affreux, mais en réalité, il n'a rien.

*Il n'y a pas eu de pluie fort dans le désert depuis le mois d'octobre dernier.

*La chorale a choisi de ne chanter que des chansons festifs au concert.

Cue + Far + Masc.

*Le col de sa chemise était autrefois blanche, mais plus maintenant.

*Le ciel dans ce tableau est lumineuse avec beaucoup de jaune.

*Le mur autour du jardin n'est pas permanente, mais sert à protéger les fleurs.

*Le plat que le chef a préparé était exquise, voilà pourquoi il a reçu des compliments.

*Le piège pour attraper des souris n'est plus bonne, par conséquent il y a des souris au sous-sol.

*Le coffre où ils ont mis leurs valises est pleine et ils sont prêts à partir.

Cue + Far + Fem

*Le météorologue a annoncé que la tempête prévue pour ce soir serait assez violent pour détruire quelques arbres.

*Le marin insiste pour que la voile de son bateau soit léger, malgré le prix.

*La serviette qu'il m'a prêtée est gris, mais très propre.

*La prise de la ville pendant la guerre étant secret, personne ne s'en est rendu compte.

*La chaise qu'il a trouvée au marche était presque gratuit, mais c'était un beau meuble ancien.

*La colline avec toutes les fleurs dessus est reposant, donc j'y vais souvent.

No cue + Far + Masc.

*L'été où le garçon a appris à nager était très chaude, plus que cet été en fait.

*Je trouve que les congés scolaires sont toujours trop brèves pour bien se détendre.

*Les poils des chats sauvages qui vivent dans les montagnes sont brunes et oranges.

*Les éclats de rire pendant la comédie étaient si bruyantes qu'on les entendait dans la rue.

*Les draps dans la chambre d'amis sont violettes et rouges.

*Les éclairs pendant une tempête sont énervantes pour tout le monde.

No cue + Far + Fem.

*Leur fierté d'avoir gagné ce match était bien apparent sur leurs visages.

*L'échelle qui se trouve dans le garage est très vieux, et pas très stable.

*Le prof se plaint que les salles de classe où il enseigne sont trop petits et souvent assez sales.

*Les deux piscines de l'hôtel qui se trouve au bord de la mer sont ouverts toute l'année.

*Les semaines qui précèdent la fin du semestre sont toujours stressants, mais se passent vite.

*Les enquêtes d'un détective privé sont souvent délicats et nécessitent de la discretion.

Filler sentences with errors

*L'orage a duré toute la nuit, mais heureusement les enfants n'ont entendu rien.

*Ne rien plait au photographe, car c'est un homme difficile.

*La montre de l'institutrice ne plus marche et elle doit en acheter une autre.

*Je n'ai pas aucune connaissance en astronomie, mais le sujet m'intéresse beaucoup.

*Il ne peut continuer plus ses études car il doit travailler à plein temps.

*Le document était compromettant, donc l'avocat a conseillé à son client de ne le divulguer pas.

*Le président est critiqué car il ne jamais répond aux questions des citoyens.

*J'ai l'impression que mon professeur n'aime pas personne dans notre cours d'histoire.

*Ma grand-mère m'est offert une armoire l'année dernière.

*Elle est marché sur la pointe des pieds pour ne pas réveiller ses parents.

*Personne n'était surpris quand ils a annoncé ses projets de déménager en Afrique.

*Les causes des manifestations actuelles ont dûes à des années de politique sociale défectueuse.

*Les deux soeurs est rentrées de l'école ensemble.

*Les timbres français est vendus uniquement dans les bureaux de tabac ou à la poste.

*Nous sommes acheté des fleurs pour sa copine qui vient d'avoir une petite fille.

*Les enfants sont choisi les mêmes jeux chaque jour pendant tout l'été.

*Après cinquante ans de travail, ils attend avec impatience leur retraite imminente.

*Les deux pays arrive enfin à avoir des rapports civils, après des années de guerre.

*Les questions qu'elle poser sont toujours d'une difficulté extrême.

*Il y a des insectes qui se dirige naturellement vers la lumière.

*Le chien veulent constamment voler la nourriture de l'assiette de son maître.

*Les trains tomber fréquemment en panne ; ce qui ennuie les passagers.

*Les filles se promenons sous le soleil qui brille.

*Les vétérinaires prend soin de tous les animaux, y compris les oiseaux et les lapins.

Correct Filler Sentences
Le propriétaire n'aime pas ces volets verts et il voudrait les peindre d'une autre couleur.
Mon camarade n'a pas pu parler à sa mère car elle était déjà sortie.
Elle n'a jamais osé dire à ses parents qu'elle avait raté son examen de français.
La vieille voiture de mon père ne tombe jamais en panne.
Nous ne voyageons jamais parce que mon mari n'aime pas les hotels.
Ce commerçant ne vend plus de papier recyclé dans son petit magasin donc il faut l'acheter ailleurs.
Le fermier n'a pas de blé cette année à cause de la sécheresse.
Il n'y a personne à la maison qui pourra aider la jeune fille avec ses devoirs de maths.
Les immeubles que cet architecte a conçus il y a 50 ans sont toujours impressionnants.
Les usines qui se trouvent dans cette ville sont polluantes.
Ce magasin a souvent des ventes exclusives pour les ses clients fidèles.
L'échec de cet homme d'affaires est honteux pour lui et aussi pour sa famille.
Après des mois de travail, elle a enfin réussi à ses examens.
Tout le monde avait remarqué l'extrême courtoisie du prince.
L'endroit où le prisonnier était retenu n'a pas été divulgué par la presse.
D'après le discours du maire, la ville sera bientôt équipée de nouveaux logements sociaux.
Pour construire son arbre généalogique, la jeune fille recherche ses racines italiennes.
Cette femme travaille souvent dans son jardin, donc elle a beaucoup de jolies fleurs.
Cet homme d'affaires a besoin de tirer de l'argent pour son voyage en Europe.
La dame a dit encore une fois à son fils qu'elle arriverait à six heures.
L'étudiant n'a pas un seul projet en tête pour son cours d'histoire.
Comme le professeur était débordé, ses collègues lui ont conseillé de faire appel à une aide extérieure.
Étant donné les événements, le capitaine s'attendait au pire.
Malgré leur valeur, ils ont dû se débarrasser de certains meubles quand ils ont déménagé.
La maman de mon voisin travaille à temps partiel chez le boulanger du coin.
La vieille dame ne sort jamais le soir.
Le médecin de mon fils habite tout près d'ici.
Quand il a vu son père déguisé en fantôme, il a vite fermé les yeux.
Les nouvelles récentes montrent que la crise économique est actuellement pire qu'avant.
Il était en prison si longtemps qu'il ne peut plus s’habituer à la vie normale.

Cette voiture se vend facilement car elle ne consomme pas beaucoup d'essence.

La relation entre le chef et son employé est meilleure que l'année dernière.

Le journal populaire de Paris a beaucoup discuté de problèmes d'immigration dans le pays.

Les séquelles de l'accident étaient suffisamment pénibles pour l'empêcher de retourner au travail.

Le président du pays a tant de décisions à prendre qu'il est très angoissé.

Les copines ont beaucoup parlé de leur séjour à la plage à leur retour.

L'avocat a admirablement plaidé le cas du pauvre mendiant.

L'écrivain veut absolument écrire un roman sur l'incident qui a eu lieu l'année dernière.

La mère a rapidement fait les courses pour la fête d'anniversaire de son fils.

L'actrice parlait passionnément de son rôle dans le dernier film qu'elle a tourné.

L'institutrice a gentiment expliqué à l'enfant qu'il n'avait pas gagné le concours.

L'athlète a spontanément répondu à l'infirmier qu'il ne s'était pas drogué.

La tragédie fait réfléchir sérieusement sur la qualité éphémère de la vie.

L'Internet va finalement faire disparaître les autres modes de communication.

Sa tante se teignait les cheveux pour paraître plus jeune.

Le touriste consultait souvent un guide, mais il s'est quand même égaré.

Elle désire consulter ses collègues avant de prendre une décision.

La jeune fille est trop influençable pour aller voir un film violent.

C'est une maison très gaie avec des grandes fenêtres et un jardin.

Cette ville n'est pas assez grande pour avoir deux théâtres.

Le locataire assure que l'appartement a été laissé en parfait état.

Il faut enseigner aux enfants à s'adresser poliment aux adultes.

L'exposition sur l'art africain aura lieu pendant la visite officielle du président du Sénégal.

Le juge va bientôt annoncer le verdict que l'on attend depuis deux semaines.

Après avoir reçu une mauvaise note, l'étudiant a décidé de travailler plus sérieusement dorénavant.

Le professeur n'a pas eu le temps de lire tout cet ouvrage.

Les acteurs étaient étrangement habillés pour le spectacle.

Ma voisine est méchante car elle maltraite mes chiens.

L'interprète parle deux langues couramment, le français et l'anglais.

Le voleur marche doucement pour ne pas faire de bruit.

Le petit garçon veut sincèrement devenir capitaine des pompiers comme son père.

L'antiquaire s'est rendu compte qu'il a fait une mauvaise affaire.

Une secrétaire devrait être discrète et efficace.

Le juge qui avait l'esprit trop occupé par ses problèmes a eu un accident de voiture.

La bibliothécaire n'achète jamais de livres car elle peut tout lire au travail.

Le boulanger a dû fermer sa boutique car il n'était pas aimable avec la clientèle.

J'ai été furieuse en rentrant de trouver toute la vaisselle sale dans l'évier.
Mes affaires sont tellement en désordre que j'ai dû engager un avocat pour m'aider.
Cette plante est si fragile qu'il faut la mettre à l'intérieur.
La maison est assez grande pour y loger toute la famille.
Les rivières sont tellement polluées qu'il est interdit d'y nager.
Son fils lui a téléphoné en arrivant pour que la famille soit rassurée.

Practice Sentences
Ce livre se vend à la librairie en ville, mais personne ne l'achète.
Le couple s'installe confortablement dans l'avion pour le long trajet à travers l'Atlantique.
Le jeune homme a hésité avant de présenter sa fiancée à sa famille.
*Ce matin mon frère me brusquement a réveillé car il voulait partir tôt.
*Si on travaille bien, on faire une bonne récolte cette année.
*L'institutrice a remercié les élèves qui lui sont envoyé des fleurs.

Table E.1

Mean (Range) for Grammaticality Judgment Task Target Nouns: Word Frequency, Word Length, Number of Syllables

Condition	Word Frequency	Word Length	Number of Syllables
Cue + Close	51.3 (21.4-127.8)	5.6 (3-8)	1.6 (1-2)
No cue + Close	42.3 (7.5-57.4)	6.5 (4-9)	1.8 (1-3)
Cue + Far	46.4 (20.7-145.0)	5.3 (3-9)	1.3 (1-2)
No cue + Far	51.62 (18.7-160.0)	6.2 (3-8)	1.8 (1-2)

Appendix F

Operation span materials

Set	Set Size	Operation	English Word	French Word	Dutch Word	Spanish Word
1	2	(18 / 3) - 4 = 2	hotel	hotel	hotel	hotel
1	2	(4 * 1) + 2 = 2	author	auteur	auteur	autor
2	2	(16 * 1) - 9 = 7	poem	poésie	gedicht	poema
2	2	(10 / 1) - 2 = 3	mouth	bouche	mond	boca
3	2	(7 * 2) - 6 = 8	piano	piano	piano	piano
3	2	(9 / 3) - 1 = 6	tree	arbre	boom	árbol
4	3	(14 * 1) - 8 = 6	group	groupe	groep	grupo
4	3	(8 / 8) + 6 = 3	rain	pluie	regen	lluvia
4	3	(20 / 2) - 9 = 1	foot	pied	voet	pie
5	3	(2 * 2) + 5 = 9	island	île	eiland	isla
5	3	(6 * 1) + 2 = 3	dust	poussière	stof	polvo
5	3	(14 / 7) + 2 = 4	clock	horloge	klok	reloj
6	3	(10 / 5) + 3 = 9	hill	colline	heuvel	libro
6	3	(5 * 2) - 5 = 9	bottle	bouteille	fles	pelo
6	3	(10 / 5) + 1 = 3	dinner	dîner	kachel	cena
7	4	(12 / 2) - 5 = 6	king	roi	koning	rey
7	4	(6 * 2) - 8 = 9	girl	fille	meisje	chica
7	4	(5 * 2) - 7 = 3	bank	banque	bank	banco
7	4	(12 / 2) - 4 = 2	lake	lac	meer	lago
8	4	(6 * 1) + 1 = 7	guide	guide	gids	guía
8	4	(20 / 5) + 5 = 5	sign	signe	teken	señal
8	4	(10 / 1) - 1 = 9	moon	lune	maan	luna
8	4	(18 * 1) - 9 = 4	bridge	pont	brug	puente
9	4	(15 * 1) - 7 = 2	knife	couteau	mes	cama
9	4	(12 / 6) + 3 = 1	chain	chaîne	ketting	queso
9	4	(20 / 4) - 3 = 2	world	monde	wereld	mundo
9	4	(3 * 1) + 1 = 4	pipe	pipe	pijp	pipa
10	5	(3 * 3) - 6 = 8	band	bande	band	banda
10	5	(3 * 2) + 3 = 9	plan	plan	plan	plan
10	5	(5 * 3) - 9 = 2	site	site	plaats	sitio
10	5	(14 / 2) - 1 = 6	leaf	feuille	blad	hoja
10	5	(14 / 7) + 1 = 9	train	train	trein	tren
11	5	(15 / 3) - 1 = 9	rifle	fusil	geweer	rifle
11	5	(7 / 7) + 1 = 2	nail	clou	spijker	clavo
11	5	(18 / 9) + 5 = 3	paper	papier	venster	papel
11	5	(8 * 2) - 9 = 3	black	noir	zwart	negro
11	5	(4 * 1) + 3 = 7	lion	lion	leeuw	león
12	5	(2 * 4) - 2 = 2	hand	main	hand	mano

12	5	(5 * 1) + 3 = 8	finger	doigt	vinger	dedo
12	5	(2 * 4) + 1 = 9	team	équipe	ploeg	flauta
12	5	(18 / 2) - 6 = 8	radio	radio	radio	radio
12	5	(16 / 2) - 5 = 3	street	rue	straat	calle
13	6	(9 / 3) + 2 = 5	valley	vallée	vallei	valle
13	6	(8 / 2) - 2 = 2	line	ligne	lijn	bota
13	6	(4 * 2) - 2 = 2	boat	bateau	boot	barco
13	6	(3 * 2) + 1 = 3	wine	vin	wijn	vino
13	6	(4 * 1) + 1 = 5	face	visage	gezicht	cara
13	6	(7 / 7) + 5 = 2	pear	poire	peer	pera
14	6	(6 * 2) - 3 = 9	rock	roche	rots	roca
14	6	(16 * 1) - 8 = 8	wall	mur	muur	pared
14	6	(9 / 3) - 2 = 1	tooth	dent	tand	diente
14	6	(15 / 3) - 4 = 1	cloud	nuage	wolk	nube
14	6	(7 * 2) - 9 = 9	floor	plancher	vloer	piso
14	6	(20 / 4) + 3 = 3	month	mois	maand	mes
15	6	(16 / 8) + 4 = 1	oven	four	oven	horno
15	6	(9 * 2) - 9 = 5	rule	règle	regel	regla
15	6	(12 / 4) + 4 = 7	beach	plage	strand	playa
15	6	(8 * 1) + 1 = 9	coast	côte	kust	costa
15	6	(2 * 3) + 2 = 4	flower	fleur	bloem	flor
15	6	(15 / 5) + 2 = 9	skirt	jupe	rok	falda

Appendix G

Language history questionnaire (English translation)

Questionnaire

This questionnaire is designed to give us a better understanding of your experience learning French. We ask that you be as accurate and thorough as possible when answering the following questions and thank you for your participation in this study.

Part I: General Information

1. Age (in years):
2. Sex: M / F
3. Handedness: L / R
4. French course(s) that you are currently enrolled in:
5. French course(s) that you are currently teaching:
6. Use of French at your job, if any:

Part II: Language History

7. What is your first or native language?
8. At what age did you first begin studying French?
9. How many years have you studied French (please include the setting(s) in which you have had experience with the language (i.e., classroom, with friends, foreign country…) and your age at each setting)

   Total number of years:

   Setting:
   Age(s):

10. Please describe any time spent in a Francophone country:

Date(s):

Length of time:

Location:

Purpose:

Living situation (i.e., host family, dorm):

Type of exposure to French (i.e., classes, work):

Percent of time spent speaking/listening/reading in French:

Other:

For the next four questions, please circle the number of your response:

9. Please rate your French reading proficiency on a ten-point scale.
(1= not literate, 10= very literate)

1 2 3 4 5 6 7 8 9 10
not literate very literate

10. Please rate your French writing proficiency on a ten-point scale.
(1= not literate, 10= very literate)

1 2 3 4 5 6 7 8 9 10
not literate very literate

11. Please rate your French conversational fluency on a ten-point scale.
(1= not fluent, 10= very fluent)

1 2 3 4 5 6 7 8 9 10
not fluent very fluent

12. Please rate your French <u>speech comprehension ability</u> on a ten-point scale.
(1= unable to understand conversation, 10= perfectly able to understand conversation)

1	2	3	4	5	6	7	8	9	10
no comprehension								perfect comprehension	

13. Have you studied any language other than French?

Yes　　　　　　　　No

If so, what languages, how old were you when you first began studying, and for how many years?

Language__________
Age started ________
Less that 2 years
2 years – 4 years
More than 4 years

Language__________
Age started ________
Less that 2 years
2 years – 4 years
More than 4 years

Language__________
Age started ________
Less that 2 years
2 years – 4 years
More than 4 years

Language__________
Age started ________
Less that 2 years
2 years – 4 years
More than 4 years

Appendix H

Experimental procedure and logistics[53]

Recruiting

The researcher contacted professors of French at universities in France and Belgium, and requested that (1) an email announcement be distributed to potential participants (professors and students of French), and (2) flyers be posted in the French departments and distributed in French classes. The email announcement and flyer contained information about the study (purpose, qualifications, payment, etc.) and the researcher's and research assistant's contact information. Potential participants contacted the researcher or research assistant directly to learn more about the study and/or sign up. In addition, an announcement was placed in an international student newsletter at a university in France. Finally, after completing the experiment, participants were given several small flyers to distribute to friends and colleagues whom they thought may be interested in participating as well.

Pre-screening

Once a potential participant expressed interest in participating, and the researcher confirmed, via email communication, that the participant was a NS of Dutch, English, or Spanish, was either enrolled in a French graduate program or working in a profession that requires use of French, and did not learn French as a child in his/her home, he/she was directed to the web-based pre-screening task.[54] The website language was French; as all

[53] Human subjects approval for all aspects of the experiment (recruitment, pre-screening, main experiment) was obtained in the standard fashion.
[54] The pre-screening website address was not included in the email announcement or on the flyers in order to allow the researcher to communicate via email or telephone with potential participants before they completed the pre-screening task. The researcher confirmed that the participants met the minimum

the participants were expected to be high level, this was not expected to pose a problem. The first page on the website described the general purpose of the experiment, and specifically, the pre-screening task (e.g., participants that fit a specific proficiency profile would be selected to participate in the main study). If participants wished to continue after reading the first page, they clicked a "continue" button located at the bottom of the page. The first page, as well as all subsequent pages, also contained a "discontinue" button that connected to a page thanking the participants for having visited the website. This allowed participants to discontinue at any time with no penalty. The second page contained the informed consent form and a check box for participants to indicate their consent to complete the pre-screening task. After checking the consent box, a blank field appeared for the participant to enter his/her name and email address in order for the researcher to contact him/her to schedule a time to complete the main experiment. After the field was filled in, a "continue" button appeared and connected to the instruction page, which explained the pre-screening task and provided an example sentence. The 14 pre-screening sentences were presented visually, one at a time. A "correct" and an "incorrect" button appeared below each sentence and the participants were instructed to decide whether the sentence was correct or incorrect as quickly as possible, without consulting external resources, such as dictionaries or the internet. Clicking the "correct" button automatically triggered the appearance of the next sentence; clicking the "incorrect" button automatically triggered a blank field to appear below the sentence, in which the participant was instructed to correct the error in the sentence. After entering the correction, participants clicked on a "continue" button to trigger the appearance of the

---

qualifications mentioned in the email announcement and on the flyer (i.e., NS of Dutch, English, or Spanish, graduate student in French) before they completed the pre-screening task.

next sentence. When the participant had judged all 14 sentences, a screen informed the participant that the researcher would contact him/her shortly to provide information about the main study. A thank you message and the researcher's contact information appeared at the bottom of the screen. The pre-screening task took approximately 10 minutes.

The pre-screening results were scored by the researcher, who then had the research assistant contact each participant to let him/her know whether he/she qualified for the main experiment, and if so, to arrange a time for the participant to come to the location of data collection.

Main Experiment

Participants were provided directions to the location in which the experiment was to take place. Each participant was tested individually by the same research assistant[55] and all tasks were administered on a laptop computer. Upon his/her arrival, the participant read and signed the informed consent form. Because participants who qualified for the main experiment had demonstrated their high-level French proficiency, the informed consent was written in French. The researcher then assigned the participant a subject number that was used to link the participant's data on each task. In addition to the subject number, the participant was randomly assigned to either Group A, Group B, or Group C, which determined which version of the gender priming task they completed in order to ensure an even distribution of participants between the three counterbalanced conditions. The participants completed a total of five tasks. The researcher opened each task file (using Psyscope), entered the subject number into the program at the start of each task,

[55] The research assistant was a NS of English, highly proficient in French. She communicated only in French with the participants.

started and stopped the digital recorder when necessary, and used a check-list for each participant, checking off the tasks as they were completed, as shown in Table H1.

Table H.1

Experimental Task Check-list

Name	#	Consent Form	GJT	Break	Gender Priming	O-Span	Gender Assign.	LHQ	Payment
John Doe	001-A	√	√	√	√	√	√	√	√
Jane Doe	002-B	√	√	√	√	√	√	√	√
Jack Doe	003-C	√	√	√	√	√	√	√	√

The participant first completed the grammaticality judgment task. The researcher sat next to the participant while he/she read through the directions and completed the practice trials to ensure the participant understood and was following the directions. The participant was able to ask questions during this time. During the experimental trials, the researcher was present, but not closely monitoring the participant. This task took approximately 30 minutes. After completing the grammaticality judgment task, participants took a 10-minute break, during which the researcher offered water and a snack.

Next, the participant completed the gender priming task. In this task, the participant saw a gender prime, followed by a target picture to be named. The Psyscope button box recorded the participant's voice onset as the RT, and a digital recorder placed next to the participant recorded the participant's responses, which were later coded for accuracy. Because this task required voice responses, the researcher closely monitored the participant throughout the task to ensure the button box microphone was picking up his/her responses. The researcher reminded the participant to speak loudly and directly into the microphone if it appeared their responses were not registering; if necessary, the researcher also reminded the participant not to say "um" or clear his/her throat before responding as these sounds affect RT measurements. This task took approximately 10 minutes.

Third, the participant completed the O-Span. The researcher sat next to the participant while he/she read through the directions and completed the practice trials to ensure the participant understood and was following the directions. If necessary, the researcher encouraged the participant to try to solve the equations more quickly (e.g., if they are timing out before the participant responded). The researcher also monitored the participant's equation accuracy on the practice trials, and, if necessary, emphasized the importance of accurately solving the equations. The participant was able to ask the researcher questions about the task instructions during this time. Once the experimental trials began, the researcher was present, but did not closely monitor the participant. This task took approximately 15 minutes.

Fourth, the participant completed the gender assignment post-test. The researcher sat next to the participant while he/she read through the directions and answered any questions, but did not monitor the participant during the task. This task took approximately 5 minutes.

Finally, the participant completed the language history questionnaire, which was presented in French in an Excel document. The researcher was present to answer any questions. This task took approximately 5 minutes. After the participant had completed all the tasks, the researcher asked if he/she had any questions about the experiment. The researcher paid the participant 15 Euros (approximately $25), and had the participant fill out the necessary payment paperwork. The entire experiment took approximately 90 minutes.[56]

[56] The entire experiment was piloted with three participants (in addition to those who piloted individual tasks) to determine whether fatigue was a factor. All three participants reported that fatigue did not affect their performance.

Appendix I

Gender priming task HLM: Model fit statistics

The following steps were followed in order to determine the fit statistics for each model.

1. Subtract the number of parameters of the previous model from the number of parameters of the new model to obtain the difference in parameters.
2. Subtract the information criteria of the first model from the information criteria of the second model in order to obtain the difference in deviation.
3. Calculate the Chi-square (χ^2) value based on the difference in parameters (alpha set at .05).
4. Compare the difference in deviation to the χ^2 critical value.

Table I.1

HLM Model Fit Statistics

	Difference in deviances	Difference in parameters (df)	χ^2 critical value	Sig.
FNS Model				
Random subject and item effects	278.821	2	5.991	$p < .05$
Main effects	25.055	8	15.51	$p < .05$
Final Interactions	53.339	57	75.62	$p > .05$; n.s.
SNS First Model				
Random subject and item effects	292.983	2	5.991	$p < .05$
Main effects	25.668	8	15.51	$p < .05$
Final Interactions	33.161	16	26.3	$p < .05$

	Difference in deviances	Difference in parameters (df)	χ^2 critical value	Sig.
SNS Second Model				
Random subject and item effects	292.983	2	5.991	$p < .05$
Main effects	35.395	11	19.68	$p < .05$
Final Interactions	74.539	39	54.57	$p < .05$
DNS First Model				
Random subject and item effects	379.094	2	5.991	$p < .05$
Main effects	32.379	8	15.51	$p < .05$
Final Interactions	60.465	36	51.0	$p < .05$
DNS Second Model				
Random subject and item effects	379.094	2	5.991	$p < .05$
Main effects	47.222	11	19.68	$p < .05$
Final Interactions	85.829	43	59.3	$p < .05$
ENS First Model				
Random subject and item effects	480.62	2	5.991	$p < .05$
Main effects	25.331	8	15.51	$p < .05$
Final Interactions	9.67	2	5.991	$p < .05$
ENS Second Model				
Random subject and item effects	480.62	2	5.991	$p < .05$
Main effects	27.586	11	19.68	$p < .05$
Final Interactions	103.351	40	55.76	$p < .05$

Appendix J

Gender priming task HLM: Complete model results

Variable	df	F-value	Sig.
FNS First Model			
Word Frequency	(1, 44)	9.865	*p* < .01
Congruency	(1, 801)	6.282	*p* < .01
Word Ambiguity	(2, 46)	3.173	n.s.
Prime Compatibility	(1, 522)	0.504	n.s.
Counterbalancing Group	(2, 21)	0.121	n.s.
SNS First Model			
Prime Compatibility x Word Ambiguity	(4, 580)	3.763	*p* < .01
Prime Compatibility x Word Ambiguity x Word Frequency	(4, 586)	3.633	*p* < .01
Congruency	(1, 868)	6.531	*p* < .05
Counterbalancing Group x Congruency	(4, 84)	3.295	*p* < .05
Word Frequency	(1, 410)	5.799	*p* < .05
Prime Compatibility x Word Frequency	(2, 603)	2.449	n.s.
Prime Compatibility	(1, 453)	2.132	n.s.
Word Ambiguity x Word Frequency	(2, 324)	1.177	n.s.

Variable	df	F-value	Sig.
Word Ambiguity	(2, 110)	1.042	n.s.
Counterbalancing Group	(2, 39)	0.316	n.s.
SNS Second Model			
Word Frequency	(1, 37)	15.116	$p < .01$
AO	(1,35)	20.923	$p < .01$
AO x Years Known	(1, 34)	15.847	$p < .01$
Years Known	(1, 34)	14.546	$p < .01$
Prime Compatibility x Word Ambiguity	(2, 808)	4.449	$p < .05$
Word Ambiguity x Years France	(2, 861)	4.484	$p < .05$
Counterbalancing Group x Years Known	(2, 34)	4.805	$p < .05$
Word Ambiguity x Congruency x Years Known	(4, 859)	3.02	$p < .05$
Prime Compatibility x Years France x Years Known	(2, 866)	3.941	$p < .05$
Counterbalancing Group x AO	(2, 35)	4.128	$p < .05$
Counterbalancing Group x Years France x Years Known	(2, 35)	4.122	$p < .05$
Counterbalancing Group	(2, 34)	4.042	$p < .05$
Prime Compatibility x Word Ambiguity x Years France	(4, 863)	2.753	$p < .05$
Counterbalancing Group x AO x Years Known	(2, 34)	3.963	$p < .05$
Counterbalancing Group x Word Ambiguity	(4, 866)	2.545	$p < .05$

Variable	df	F-value	Sig.
Prime Compatibility x Years France	(2, 869)	3.179	$p < .05$
Years France	(1, 39)	2.569	n.s.
Prime Compatibility x Years Known	(1, 857)	1.764	n.s.
Word Ambiguity x Years Known	(2, 857)	1.617	n.s.
Word Ambiguity	(2, 120)	1.483	n.s.
Prime Compatibility	(1, 894)	1.262	n.s.
Word Ambiguity x Congruency	(2, 857)	1.003	n.s.
Congruency x Years Known	(1, 860)	0.391	n.s.
Years France x Years Known	(1, 37)	0.289	n.s.
Congruency	(1, 855)	0.251	n.s.
DNS First Model			
Counterbalancing Group x Word Frequency	(2, 754)	5.656	$p < .01$
Counterbalancing Group x Word Ambiguity x Word Frequency	(4, 795)	3.735	$p < .01$
Prime Compatibility x Word Ambiguity x Word Frequency	(4, 389)	3.463	$p < .01$
Word Frequency	(1, 191)	6.288	$p < .05$
Counterbalancing Group x Word Ambiguity x Congruency	(8, 107)	2.491	$p < .05$
Prime Compatibility x Word Ambiguity	(2, 264)	3.378	$p < .05$
Counterbalancing Group x Congruency	(4, 106)	2.418	n.s.

Variable	df	F-value	Sig.
Word Ambiguity x Word Frequency	(2, 192)	2.282	n.s.
Counterbalancing Group	(2, 57)	2.296	n.s.
Congruency	(1, 899)	2.531	n.s.
Prime Compatibility x Word Frequency	(2, 349)	1.906	n.s.
Word Ambiguity	(2, 115)	1.742	n.s.
Prime Compatibility	(1, 228)	0.979	n.s.
Word Ambiguity x Congruency	(2, 897)	1.008	n.s.
Counterbalancing Group x Word Ambiguity	(4, 922)	1.03	n.s.
DNS Second Model			
Word Frequency	(1, 69)	68.89	$p < .01$
Counterbalancing Group x Congruency	(4, 339)	339.391	$p < .01$
Word Frequency x AO x Years Known	(1, 885)	884.63	$p < .01$
Congruency x Years France	(2, 887)	887.029	$p < .01$
Counterbalancing Group x Congruency x Years Known	(4, 891)	891.032	$p < .01$
Congruency	(1, 891)	891.048	$p < .01$
Word Ambiguity x AO x Years Known	(2, 890)	889.869	$p < .01$
Word Ambiguity x Years Known	(2, 890)	890.211	$p < .01$
Congruency x AO	(2, 893)	892.761	$p < .01$

Variable	df	F-value	Sig.
Counterbalancing Group x Word Ambiguity x Years France	(4, 888)	887.785	$p < .01$
Congruency x Years France x Years Known	(2, 885)	884.831	$p < .01$
Word Ambiguity x AO	(2, 889)	888.967	$p < .05$
Word Ambiguity	(2, 893)	893.435	$p < .05$
Congruency x Years Known	(2, 888)	887.821	$p < .05$
Counterbalancing Group x Word Ambiguity x Years Known	(4, 889)	889.028	$p < .05$
Counterbalancing Group x Word Ambiguity	(4, 889)	888.633	n.s.
Years France x Years Known	(1, 34)	33.671	n.s.
Years France	(1, 35)	34.769	n.s.
Counterbalancing Group	(2, 36)	36.26	n.s.
Counterbalancing Group x Years Known	(2, 36)	35.993	n.s.
AO x Years Known	(1, 36)	36.46	n.s.
Years Known	(1, 36)	36.28	n.s.
Counterbalancing Group x Years France	(2, 35)	35.218	n.s.
Prime Compatibility	(1, 587)	586.654	n.s.
AO	(1, 37)	37.187	n.s.

Variable	df	F-value	Sig.
ENS First Model			
Word Frequency	(1, 42)	15.281	$p < .01$
Counterbalancing Group x Congruency	(4, 287)	3.164	$p < .05$
Prime Compatibility	(1, 1080)	3.997	$p < .05$
Congruency	(1, 1232)	3.049	n.s.
Counterbalancing Group	(2, 41)	0.072	n.s.
ENS Second Model			
Counterbalancing Group	(2, 46)	14.666	$p < .01$
Word Frequency	(1, 60)	21.218	$p < .01$
Counterbalancing Group x AO	(2, 45)	15.374	$p < .01$
AO x Years Known	(1, 41)	16.679	$p < .01$
Years France	(1, 42)	13.143	$p < .01$
Years Known	(1, 42)	13.261	$p < .01$
AO x Years France	(1, 41)	13.763	$p < .01$
Word Frequency x AO x Years Known	(1, 1219)	11.091	$p < .01$
Counterbalancing Group x Years France	(2, 41)	7.004	$p < .01$
Counterbalancing Group x Congruency x Years France	(4, 1218)	4.207	$p < .01$
Counterbalancing Group x AO x Years Known	(2, 41)	7.638	$p < .01$

Variable	df	F-value	Sig.
Counterbalancing Group x Years Known	(2, 42)	6.904	$p < .01$
Prime Compatibility x Years France	(1, 1226)	8.254	$p < .01$
Counterbalancing Group x Congruency x Years Known	(4, 1225)	3.927	$p < .01$
Counterbalancing Group x Years France x Years Known	(2, 41)	6.411	$p < .01$
Counterbalancing Group x AO x Years France	(2, 41)	5.991	$p < .01$
Years France x Years Known	(1, 42)	7.883	$p < .01$
Counterbalancing Group x Congruency	(4, 1254)	3.131	$p < .05$
Word Frequency x Years France x Years Known	(1, 1218)	5.749	$p < .05$
Counterbalancing Group x Congruency x AO	(6, 1220)	2.28	$p < .05$
Congruency x Years France	(1, 1217)	3.671	n.s.
Congruency x Years Known	(2, 1221)	1.402	n.s.
Word Ambiguity	(2, 46)	1.062	n.s.
Congruency	(1, 1231)	0.495	n.s.
AO	(1, 45)	0.057	n.s.
Prime Compatibility	(1, 1234)	0.039	n.s.

Appendix K

Gender priming task HLM: Interactions

This appendix presents the two- and three-way interactions that were significant in the HLM models but are not relevant to gender priming effects. The primary goal of the HLM models was to determine whether the participants demonstrate gender priming effects, and whether additional factors, such as AO, number of years spent in France, and number of years a participant has know French, predict these effects. The interactions between variables that do not include Congruency are discussed here, as they do not help to explain the presence or absence of gender priming effects, and would be disruptive to the flow of the discussion in the main text. Furthermore, in many cases it is difficult to interpret these interactions because one or more of the variables represents a continuous variable, and/or there are a small number of subjects and/or items per cell. Therefore, although the findings are presented here, it is important not to over interpret the results. In addition, the interaction between Counterbalancing Group and Congruency is considered for each language group in order to examine the potential role of list and/or group effects in this task.

SNS Interactions

Although the SNSs showed a main effect of Congruency in the first HLM model, it was subsumed by a significant two-way interaction between Counterbalancing Group and Congruency, raising the question as to whether list and/or group effects are driving the congruency effect. Figure 9 in Section 8.4.2, repeated here as Figure K.1 for the reader's convenience, confirms that participants were fastest at naming pictures in List 3, and slowest at naming pictures in List 1, regardless of congruency condition; therefore, the interaction between Counterbalancing Group and Congruency may be explained by list effects rather than by a priming effect that occurs only for Group C.

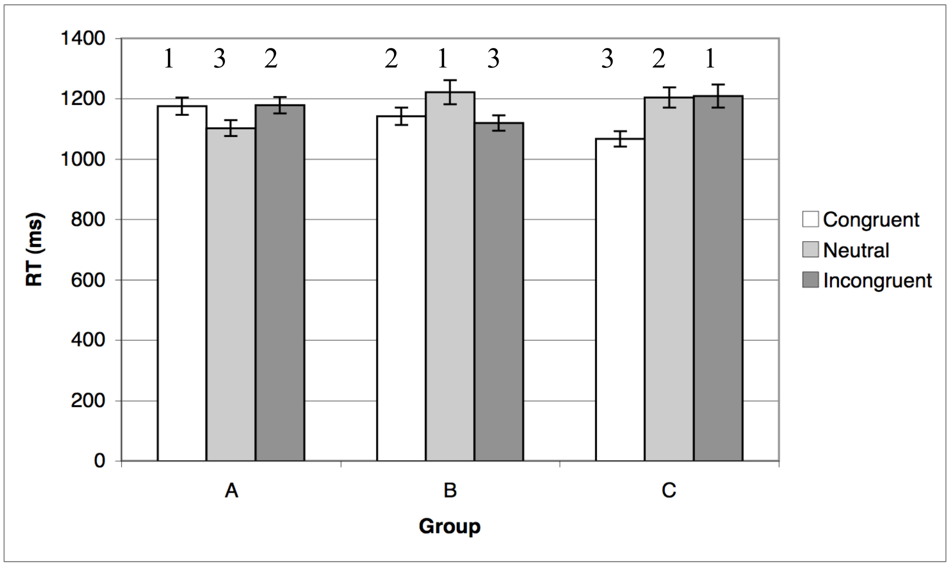


Figure K.1. SNS Counterbalancing Group by Congruency interaction, with list (1, 2, 3) displayed above each column to indicate which Group saw which list in each Congruency condition

The role of list effects for the DNS and ENS participants are addressed in the DNS and ENS sections of this appendix, but it is worth noting here that all three NNS groups showed the fastest overall RTs for List 3 and the slowest overall RTs for List 1. Interestingly, the FNSs did not show the same pattern; List 2 and List 3 RTs were similar to each other, and both faster than List 1 RTs. The mean RTs for the three lists for each language group are presented in Table K.1.

Table K.1

Mean RT (ms) for Lists 1, 2, and 3 for Each Language Group

	List 1	List 2	List 3
FNS	915	854	857
SNS	1202	1175	1096
DNS	1178	1157	1117
ENS	1163	1116	1044

That the pattern of list RTs is consistent across NNS groups suggests there is some characteristic of the lists that is driving the SNS Counterbalancing Group by Congruency interaction. However, when the lists were compared for word frequency, name agreement, image agreement, image familiarity, and image complexity, the number of nouns with an unvoiced onset, and the number of vowel-initial nouns, as shown in Table K.2, no characteristic, thus far, stands out as a potential factor that could explain the faster RTs for List 3.

Table K.2

Characteristics of Lists 1, 2, and 3

	List 1	List 2	List 3
Word frequency (per million)	58.2 (1.4-479.9)	57.7 (2.9-504.3)	54.8 (7.6-306.5)
Name agreement	99.1 (93-100)	99.8 (96-100)	98.8 (93-100)
Image agreement	3.7 (2.4-4.6)	3.7 (2.9-4.5)	3.3 (2.1-4.7)
Image familiarity	2.9 (1.8-4.8)	4.0 (2.1-5.0)	3.6 (1.5-5.0)
Image complexity	2.8 (1-4.9)	2.7 (1.6-4.6)	2.7 (1.2-4.4)
Number (%) of words with unvoiced onset	9 of 16 (56%)	8 of 16 (50%)	10 of 16 (62%)
Number (%) of vowel-initial words	3 of 16 (19%)	2 of 16 (13%)	2 of 16 (13%)

Furthermore, it is important to note that the FNSs, who saw the same lists and were assigned to groups in the same manner, showed a main effect of Congruency, with no Counterbalancing Group or list effects.

Another possible explanation for the Counterbalancing Group by Congruency interaction is that there exists a difference among the three groups of SNS participants. Table K.3 presents the language history questionnaire data and performance on three language tasks (grammaticality judgment filler sentence accuracy, gender priming picture naming accuracy, and gender assignment post-test accuracy) for the three groups.

Table K.3

SNS Language History Questionnaire data and Language Task Performance

	Group A	Group B	Group C
Age	29.8 (24-38)	28.7 (21-45)	27.5 (20-41)
AO	17.8 (12-31)	16.4 (10-27)	18.2 (10-30)
Years France	3.4 (.3-8.5)	3.6 (.5-14)	2.1 (.2-4.8)
Years Known	12.0 (3-20)	12.3 (4-23)	9.3 (2-18)
Reading	8.6 (8-10)	8.4 (7-10)	8.0 (6-10)
Writing	7.3 (5-10)	7.3 (4-9)	6.0 (3-8)
Speaking	7.6 (6-9)	7.7 (5-9)	7.2 (4-9)
Comprehension	8.9 (7-10)	8.9 (6-10)	8.4 (4-10)
Grammaticality judgment filler sentences	86% (77-93%)	87% (74-93%)	84% (75-93%
Gender priming picture naming	63% (44-77%)	67% (42-85%)	60% (46-73%
Gender assignment post-test	92% (85-99%)	92% (81-100%)	85% (74-97%)

A series of one-way ANOVAs showed that the three groups differed significantly as a function of gender assignment accuracy, $F(2, 34) = 5.219$, $p < .05$, and self-reported writing proficiency, $F(2, 34) = 3.760$, $p < .05$. Tukey post-hoc comparisons showed that Group C was significantly lower than both Groups A and B for gender assignment accuracy ($p < .05$ for both comparisons) and significantly lower than Group B for writing ($p < .05$); but Groups A and B did not differ significantly from each other. Because priming effects indicate native-like gender representation, a higher level of proficiency might explain potential priming effects for Group C. However, the lower gender assignment accuracy and writing self-ratings for Group C indicate a lower level of proficiency compared to Groups A and B, and, therefore, neither of these differences explain the potential priming effects for Group C.

A three-way interaction between Prime Compatibility, Word Ambiguity, and Word Frequency was found in the first SNS HLM analysis. Because Word Frequency constitutes a continuous variable, it was divided into three categorical groups (low frequency = 1-12 per million, $n = 17$; medium frequency = 13-33 per million, $n = 17$; high frequency = 36-504 per million, $n = 14$) for the purpose of visually representing the data. The divisions were made on the basis of creating equal groups in order to minimize the number of empty cells for Prime Compatibility and Word Ambiguity. The results are presented in Figures K.2-K.4.

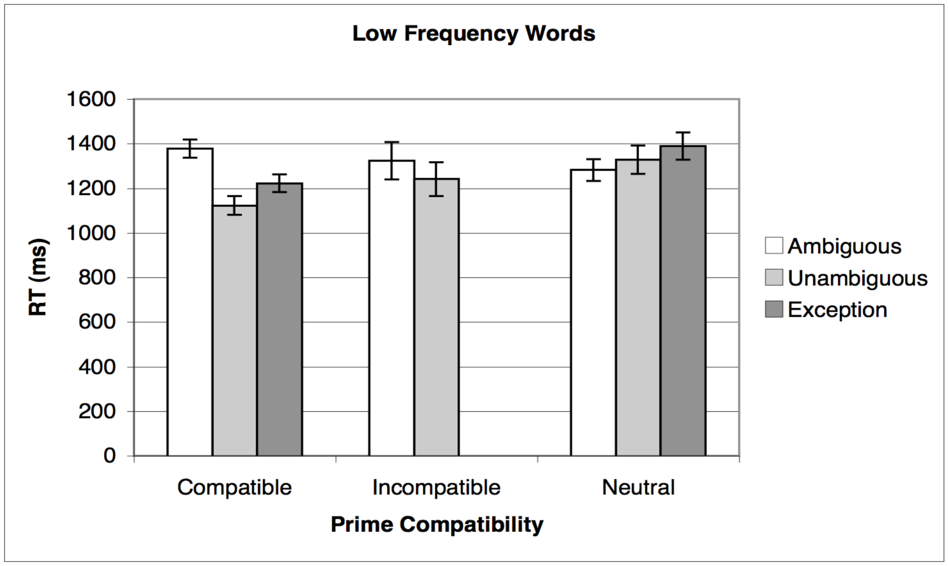


Figure K.2. SNS interaction between Prime Compatibility and Word Ambiguity: low frequency words

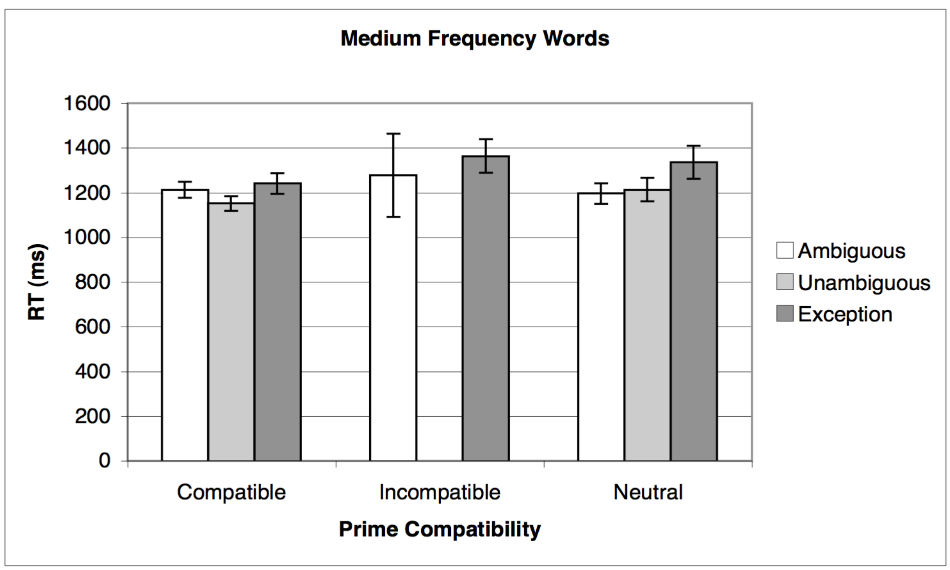


Figure K.3. SNS interaction between Prime Compatibility and Word Ambiguity: medium frequency words

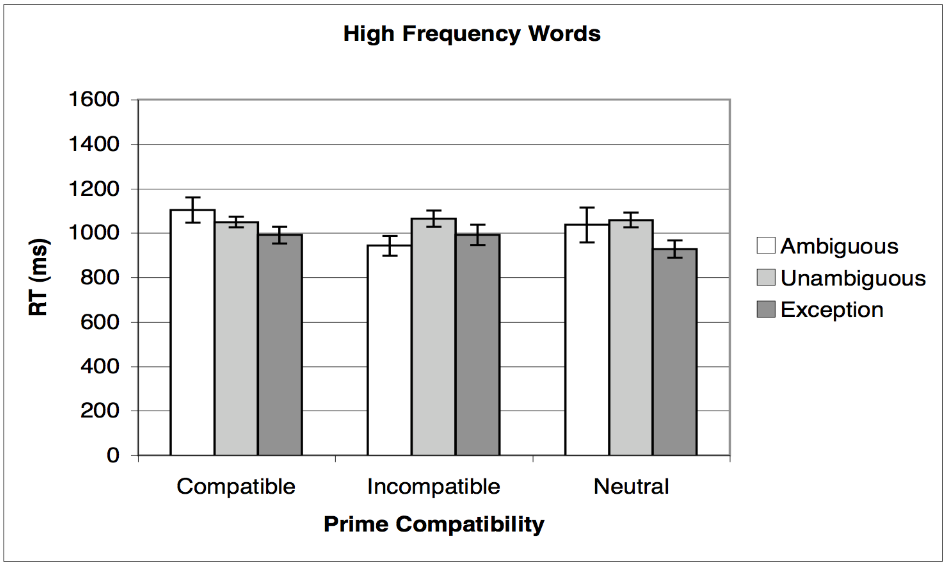


Figure K.4. SNS interaction between Prime Compatibility and Word Ambiguity: high frequency words

For low frequency words matched with compatible primes, participants showed fastest RTs in the unambiguous condition and slowest RTs in the ambiguous condition, a finding that mirrors the main effect of Word Ambiguity among FNSs. However, for the SNSs, this pattern only occurs when the prime is compatible and diminishes with medium and high frequency words. No pattern emerges for low frequency words with incompatible and neutral primes. For medium frequency words, RTs in the exception condition were slowest regardless of prime compatibility, but for high frequency words, RTs in the exception condition were fastest with compatible and neutral primes. Also for high frequency words, RTs in the ambiguous condition were slowest with compatible primes but fastest with incompatible primes and RTs in the unambiguous condition were slowest with incompatible primes. Overall, the only potential pattern that emerges is faster RTs in the unambiguous condition and slower RTs in the ambiguous condition for low and medium frequency words when the prime is compatible. This sensitivity to Word Ambiguity will be addressed in Section 8.4.5 of the main text.

A three-way interaction between Prime Compatibility, Word Ambiguity, Years France was found to be significant in the second analysis (Figures K.5-K.7).

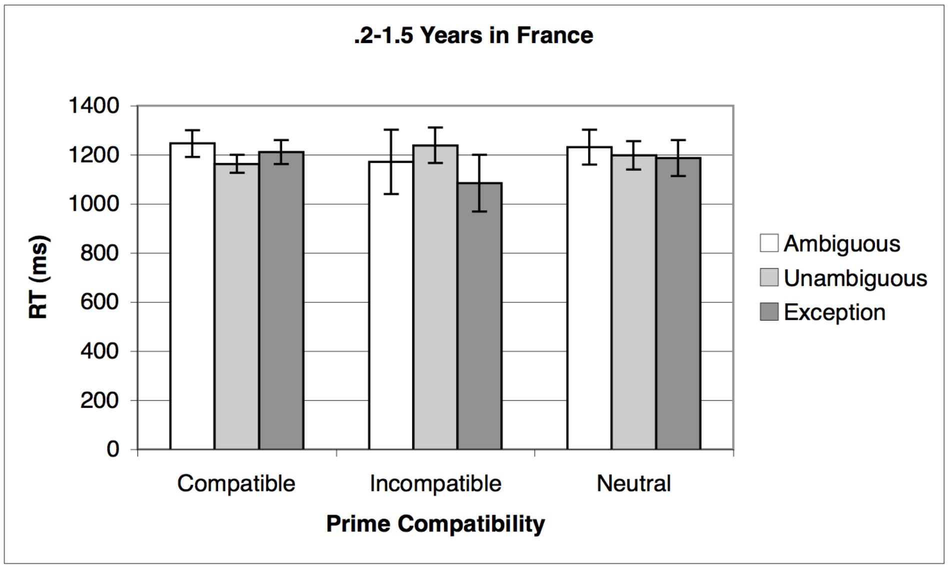


Figure K.5. SNS interaction between Prime Compatibility and Word Ambiguity for participants who had spent .2-1.5 years in France

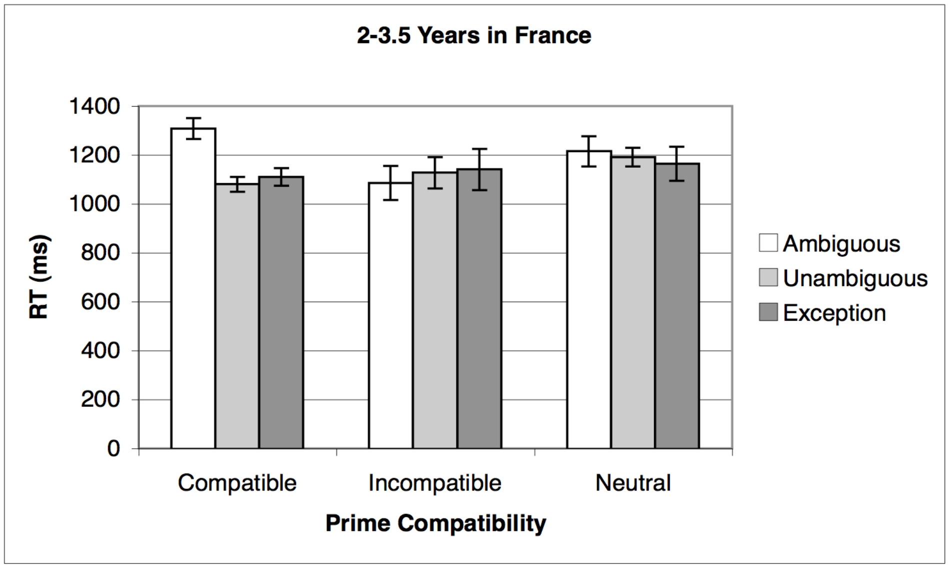


Figure K.6. SNS interaction between Prime Compatibility and Word Ambiguity for participants who had spent 2- 3.5 years in France

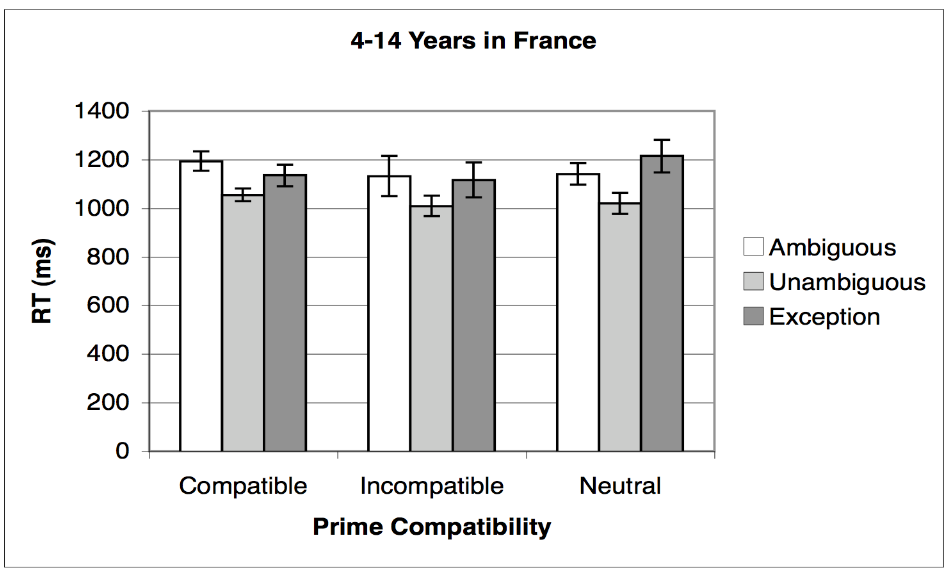


Figure K.7. SNS interaction between Prime Compatibility and Word Ambiguity for participants who had spent 4-14 years in France

For the participants who had spent the least amount of time in France (.2-1.5 years), RTs were fastest for unambiguous nouns when the prime was compatible, but similar among ambiguous, unambiguous, exception nouns when the prime was incompatible or neutral. For the participants who had spent 2-3.5 years in France, the only clear difference between Word Ambiguity is the slower RTs for ambiguous nouns when the prime was compatible, and for the participants who had spent the most time in France (4-14 years), RTs were faster for unambiguous nouns, regardless of the prime compatibility. Again, a sensitivity to Word Ambiguity similar to that found with the FNSs is evident. Generally, RTs were faster for unambiguous nouns and slower for ambiguous nouns. However, this pattern is not consistent across compatible, incompatible, or exception words, or the number of years a participant had spent in France.

A significant two-way interaction between Counterbalancing Group and Word Ambiguity was also found. As shown in Figure K.8, participants randomly assigned to Group B showed faster RTs for unambiguous nouns as compared to similar RTs for ambiguous and exception nouns, where as participants assigned to Group C showed slower RTs for ambiguous nouns as compared to similar RTs for unambiguous and exception nouns. The difference between Word Ambiguity RTs for Group A are minimal. Once again, although a pattern of faster RTs for unambiguous nouns as compared to ambiguous nouns is apparent, it does not occur for all participants.

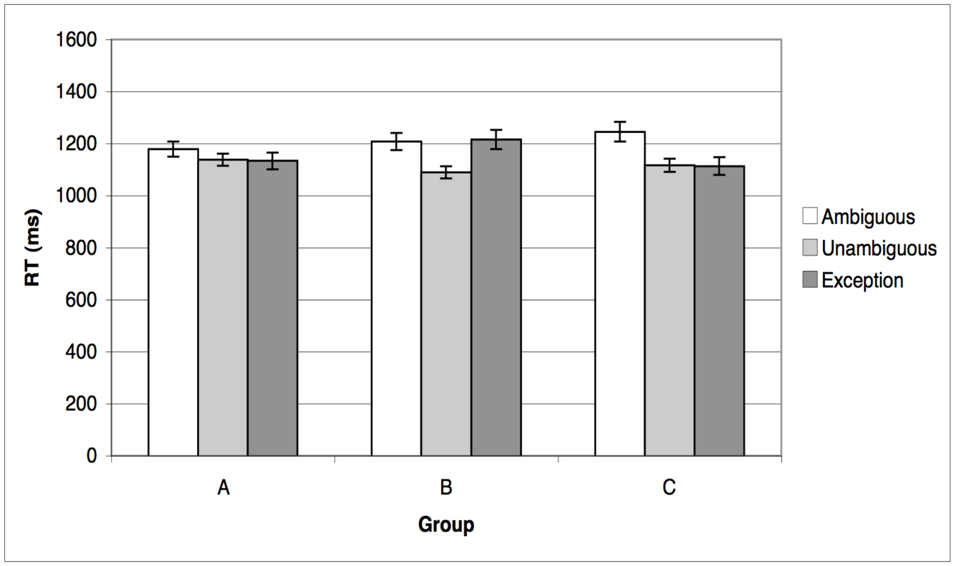


Figure K.8. SNS interaction between Counterbalancing Group and Word Ambiguity

Finally, the following three-way interactions were all significant in the second HLM analysis for SNSs:

- Prime Compatibility, Years France, Years Known
- Counterbalancing Group, Years France, Years Known
- Counterbalancing Group, AO, Years Known

However, due to the small number of subjects per cell, and in several cases, empty cells, as a result of creating categorical variables out of continuous variables, it is not possible to interpret these results, nor would an interpretation be meaningful or relevant to the goal of the analysis. Therefore, these interactions are not addressed.

DNS Interactions

Figure K.9 shows the interaction between Counterbalancing Group and Congruency, along with the lists each group saw. Similar to the SNS interaction, picture naming times are always fastest for List 3, regardless of the congruency condition.

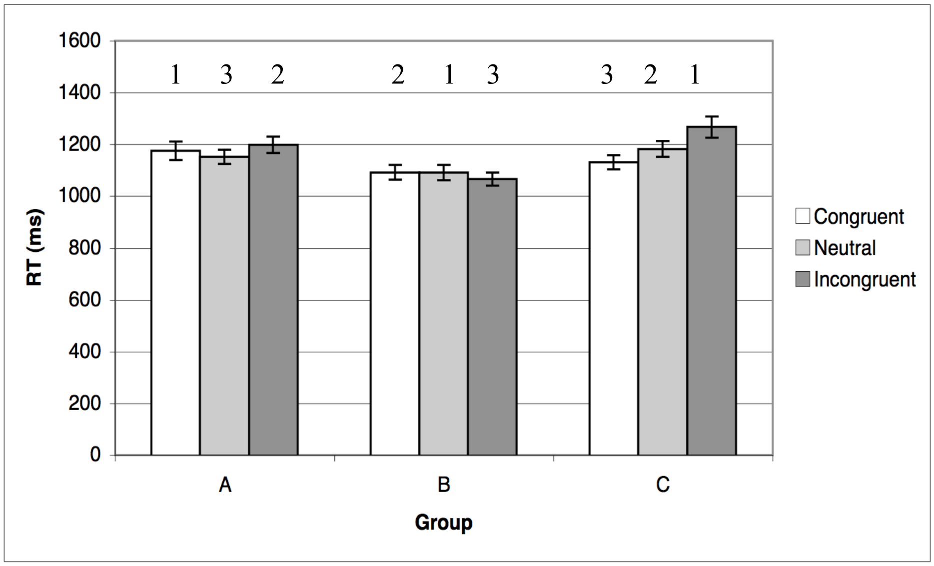


Figure K.9. DNS Counterbalancing Group by Congruency interaction, with list (1, 2, 3) displayed above each column to indicate which Group saw which list in each Congruency condition

Also, in considering any patterns including Counterbalancing Group, it is important to note that a series of one-way ANOVAs comparing the three DNS groups on language history questionnaire data and accuracy on grammaticality judgment filler sentences, gender priming picture naming, and gender assignment (displayed in Table K.4) revealed no significant differences between any of the Counterbalancing Groups on any of the factors.

Table K.4

DNS Language History Questionnaire data and Language Task Performance

	Group A	Group B	Group C
Age	26.2 (20-46)	31.0 (20-61)	26.7 (20-43)
AO	10.9 (10-13)	10.6 (8-13)	10.3 (9-12)
Years France	3.1 (.8-24)	7.2 (.1-38)	1.6 (.08-8)
Years Known	15.3 (10-36)	20.4 (10-50)	16.3 (10-32)
Reading	8.0 (5-10)	8.8 (7-10)	8.6 (8-9)
Writing	7.4 (6-9)	7.4 (5-10)	7.4 (5-9)
Speaking	7.5 (6-9)	7.7 (4-10)	7.2 (5-10)
Comprehension	8.8 (7-10)	8.9 (4-10)	8.6 (7-10)
Grammaticality judgment filler sentences	87% (80-95%)	91% (77-99%)	89% (77-97%)
Gender priming picture naming	61% (46-90%)	71% (42-94%)	60% (44-83%)
Gender assignment post-test	91% (80-97%)	93% (86-98%)	93% (69-100%)

A three-way interaction between Counterbalancing Group, Word Ambiguity, and Word Frequency was found and RTs are displayed in Figures K.10-K.12. The same categorical groups that were created for Word Frequency in the SNS analysis were used for the DNSs, with low frequency words ranging from 1-12 per million, medium frequency words ranging from 12-33 million, and high frequency words ranging from 36-504 million.

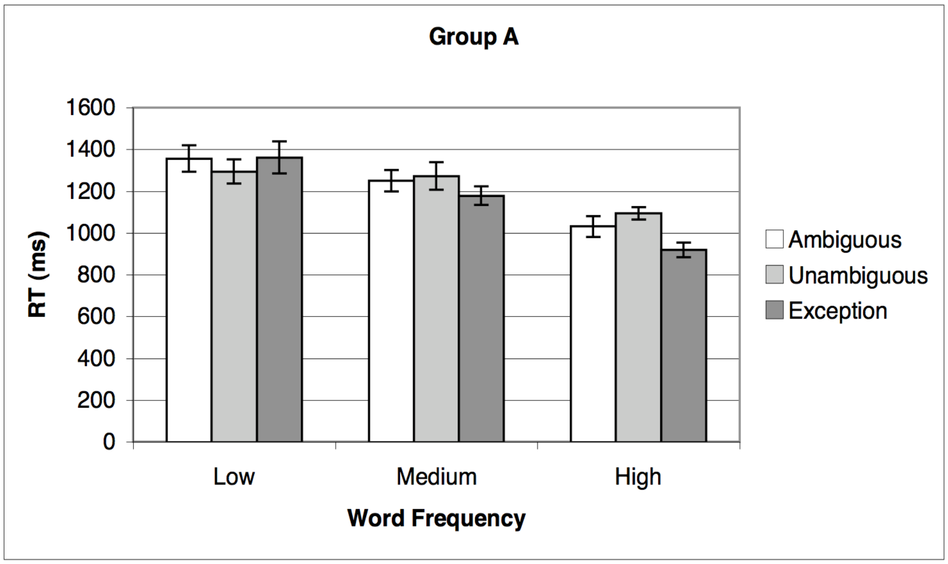


Figure K.10. DNS interaction between Word Ambiguity and Word Frequency: Group A

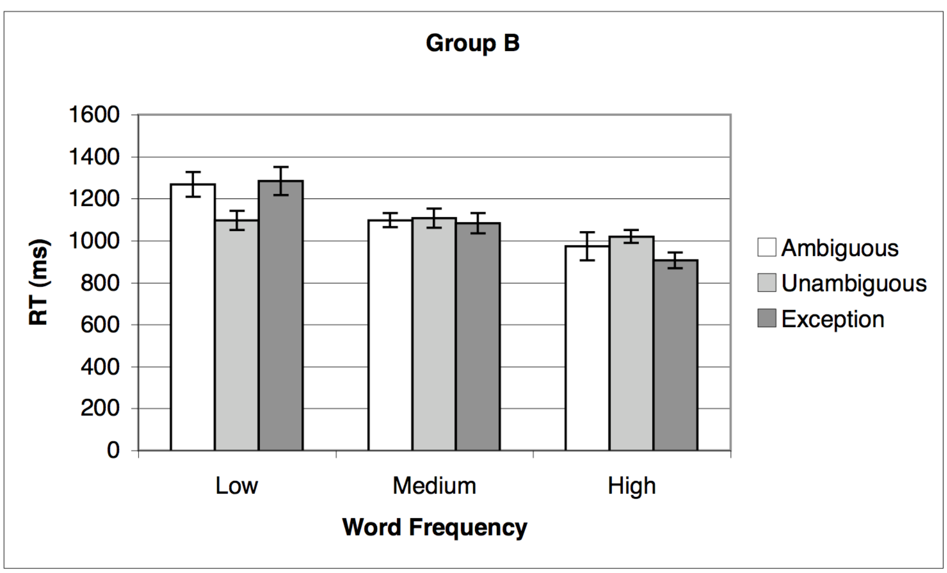


Figure K.11. DNS interaction between Word Ambiguity and Word Frequency: Group B

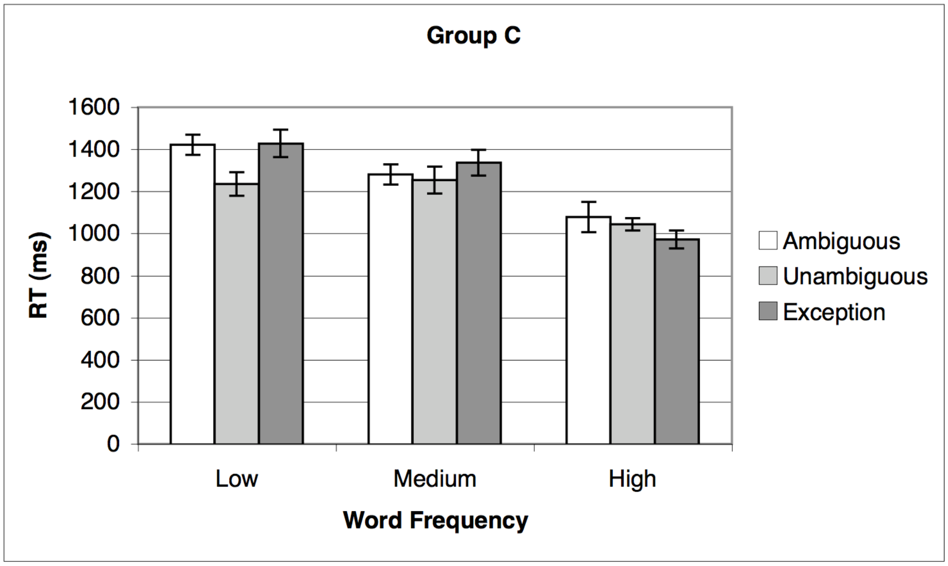


Figure K.12. DNS interaction between Word Ambiguity and Word Frequency: Group C

Participants in all three groups showed faster RTs for high frequency words as compared to low frequency words. Participants in Groups B and C, but not A, showed faster RTs for low frequency unambiguous nouns than for low frequency ambiguous and exception nouns. For medium frequency words, none of the three groups showed differences in RT as a function of Word Ambiguity. All three groups showed faster RTs for high frequency exception nouns, as compared to high frequency ambiguous and unambiguous nouns. Overall, the DNSs demonstrate a sensitivity to Word Ambiguity, although this sensitivity is not consistent across Counterbalancing Group or Word Frequency.

A three-way interaction between Prime Compatibility (syntactically compatible vs. incompatible prime-target combinations), Word Ambiguity, and Word Frequency was also found. This interaction is represented in Figures K.13-K.15; however, the small number of items and two empty cells make this interaction difficult to interpret.

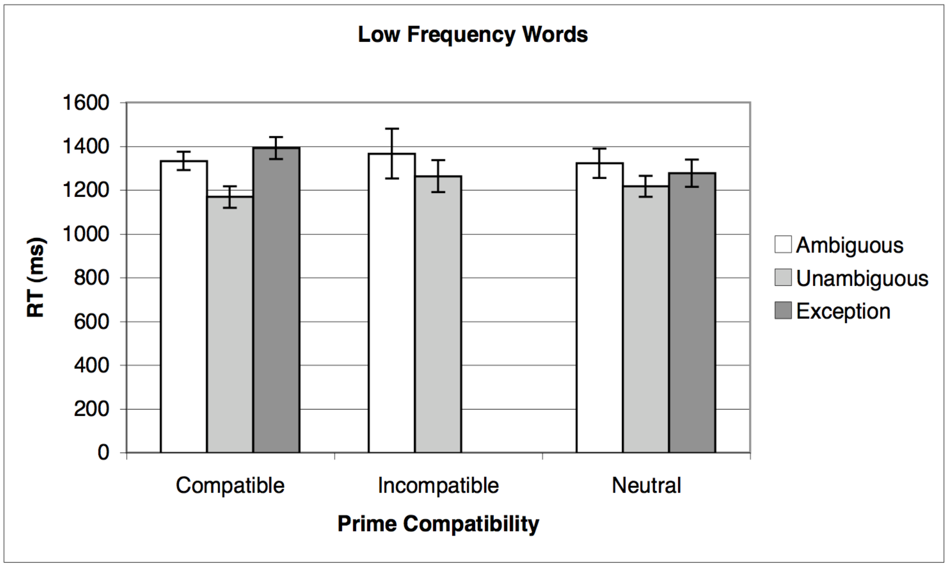


Figure K.13. DNS interaction between Prime Compatibility and Word Ambiguity: low frequency words

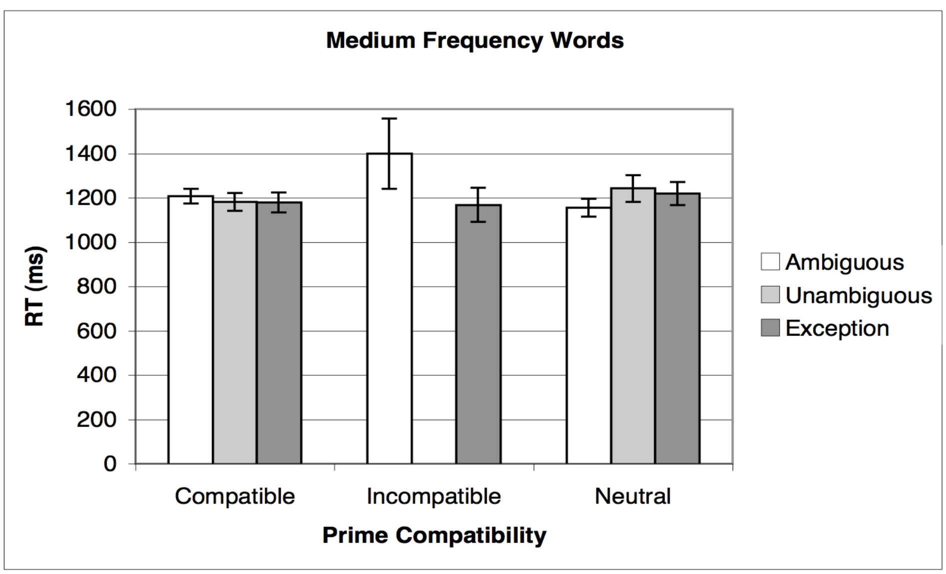


Figure K.14. DNS interaction between Prime Compatibility and Word Ambiguity: medium frequency words

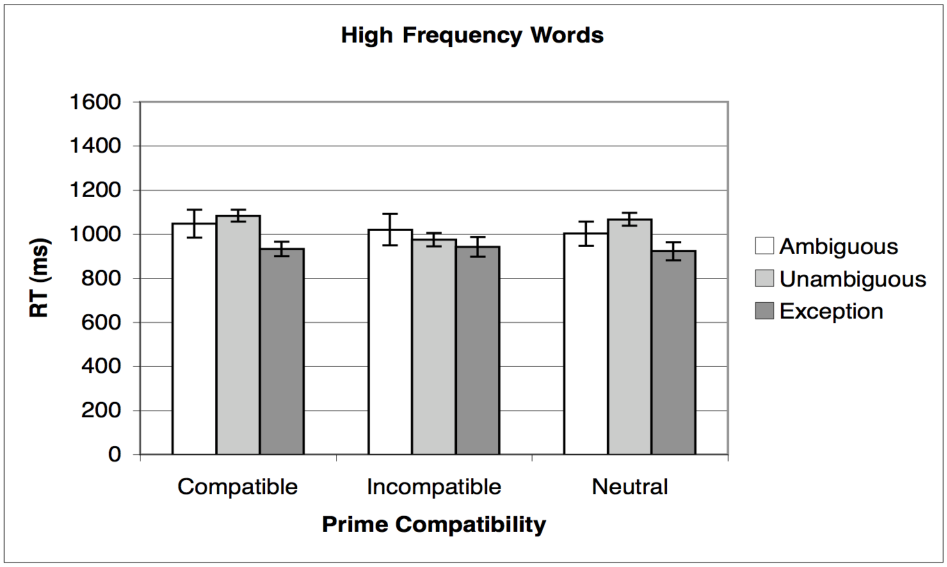


Figure K.15. DNS interaction between Prime Compatibility and Word Ambiguity: high frequency words

For low frequency words, RTs for unambiguous nouns were fastest in the compatible and neutral conditions, but this effect disappeared for medium frequency words, with similar RTs across the ambiguity conditions with compatible primes, slightly faster RTs for ambiguous nouns in the neutral condition, and not enough data points for comparison in the incompatible condition. For high frequency words, exception noun RTs were the fastest regardless of the prime. Overall, no clear picture naming RT pattern emerges from this interaction. At most, RTs were faster for low frequency, unambiguous nouns when the prime is compatible or neutral, which is similar to the pattern seen with SNSs. But this pattern shifts to faster RTs for exception nouns with compatible and neutral primes when the nouns are high frequency. The DNS sensitivity to Word Ambiguity in picture naming RTs will be addressed in Section 8.4.5 of the main text.

Finally, the following three-way interactions were all significant in the second HLM analysis for DNSs:

- Counterbalancing Group, Word Ambiguity, Years France
- Counterbalancing Group, Word Ambiguity, Years Known
- Word frequency, AO, Years Known
- Word Ambiguity, AO, Years Known

However, due to the small number of subjects or items per cell, and in several cases, empty cells, as a result of creating categorical variables out of continuous variables, it is not possible to interpret these results, nor would an interpretation be meaningful or relevant to the goal of the analysis. Therefore, these interactions are not addressed.

ENS Interactions

The significant two-way interaction between Counterbalancing Group and Congruency indicates that participants showed differing priming effects based on their Counterbalancing Group. However, as with the SNSs, this effect is likely due to list effects, as shown in Figure K.16.

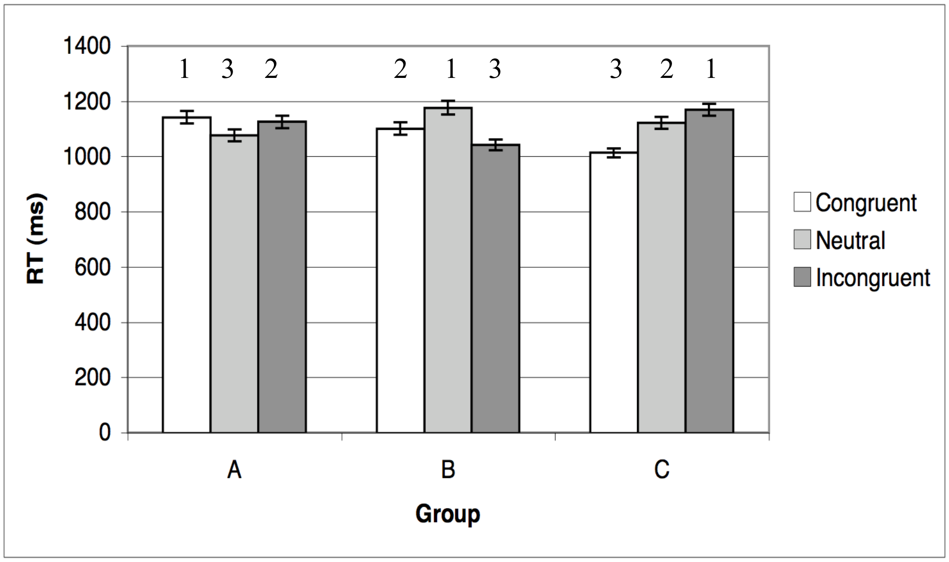


Figure K.16. ENS Counterbalancing Group by Congruency interaction, with list (1, 2, 3) displayed above each column to indicate which Group saw which list in each Congruency condition

To verify that there were no differences between the three groups, language history questionnaire data and performance on three language tasks (grammaticality judgment filler sentence accuracy, gender priming picture naming accuracy, and gender assignment post-test accuracy) were compared, as shown in Table K.5. A series of one-way ANOVAs confirmed that there were no significant differences among any of the groups for any of the factors. In other words, there are no apparent characteristics of the groups that could explain the potential priming effects for Group C.

Table K.5

ENS Language History Questionnaire data and Language Task Performance

	Group A	Group B	Group C
Age	45.8 (20-65)	39.1 (23-59)	45.3 (23-67)
AO	16.5 (9-38)	14.0 (11-23)	15.4 (11-21)
Years France	20.3 (1-41)	11.7 (.75-37)	14.3 (.6-37)
Years Known	29.3 (4-53)	25.1 (6-47)	29.3 (6-55)
Reading	8.5 (7-10)	8.2 (7-10)	8.1 (4-10)
Writing	6.9 (3-9)	6.7 (4-9)	6.4 (2-9)
Speaking	7.7 (6-9)	7.7 (5-9)	7.3 (4-10)
Comprehension	8.8 (6-10)	9.0 (7-10)	8.5 (4-10)
Grammaticality judgment filler sentences	88% (73-95%)	87% (71-98%)	83% (58-96%)
Gender priming picture naming	79% (54-92%)	76% (48-98%)	75% (52-90%)
Gender assignment post-test	89% (70-97%)	90% (78-96%)	88% (79-96%)

As with the SNSs, it is impossible to determine how much of the interaction is due to list effects and how much is due to differences across the three groups. Therefore, the data showing faster RTs in the congruent condition than in the incongruent condition for Group C provides, at best, very weak evidence for priming.

The following three-way interactions were all significant in the second HLM analysis for ENSs:

- Counterbalancing Group, AO, Years Known
- Counterbalancing Group, AO, Years France
- Group, Years France, Years Known
- Word Frequency, AO, Years Known
- Word Frequency, Years France, Years Known

However, due to the small number of subjects per cell, and in several cases, empty cells, as a result of creating categorical variables out of continuous variables, it is not possible to interpret these results, nor would an interpretation be meaningful or relevant to the goal of the analysis. Therefore, these interactions are not addressed.

Bibliography

Alario, F. X., & Caramazza, A. (2002). The production of determiners: Evidence from French. *Cognition, 82,* 179-223.

Alario, F. X., & Ferrand, L. (1999). A set of 400 pictures standardized for French: Norms for name agreement, image agreement, familiarity, visual complexity, and age of acquisition. *Behavior Research Methods, Instruments, & Computers, 31*, 531-552.

Alario, F. X., Matos, R. E., & Segui, J. (2004). Gender congruency effects in picture naming. *Acta Psychologica, 117,* 185-204.

Arnaud, P. J. L. (1999). Target-error resemblance in French word substitution speech errors and the mental lexicon. *Applied Psycholinguistics, 20,* 269-287.

Augé, Gillon, Hollier-Larousse, Moreau, & Cie (1959). *Petit Larousse* Paris: Librairie Larousse.

Bartning, I. (2000). Gender agreement in L2 French: Pre-advanced vs. advanced learners. *Studia Linguistica, 54*, 225-237.

Bartning, I., & Schlyter, S. (2004). Itinéraires acquisitionnels et stades de développement en français L2. *French Language Studies, 14,* 281-299.

Bialystok, E. (1997). The structure of age: in search of barriers to second language acquisition. *Second Language Research, 13*, 116-137.

Blackwell, A., Bates, E., & Fisher, D. (1996). The time course of grammaticality judgment. *Language and Cognitive Processes, 11*, 337-406.

Blom, E., Polišenská, D., & Unsworth, S. (2008). The acquisition of grammatical gender in Dutch. *Second Language Research* 24, 259-265.

Bordag, D., Opitz, A., & Pechmann, T. (2006). Gender processing in first and second languages: The role of noun termination. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32,* 1090-1101.

Caramazza, A., Miozzo, M., Costa, A., Schiller, N. o., & Alario, F. X. (2001). A cross-linguistic investigation of determiner production. In E. Dupoux (Ed.), *Language, brain, and cognitive development: Essays in honor of Jaques Mehler* (pp. 209-226). Cambridge, MA: MIT Press.

Carroll, S. (1989). Second-language acquisition and the computational paradigm. *Language Learning, 39,* 535-594.

Clarke, E. (1985). The acquisition of Romance, with special reference to French. In D. I. Slobin (Ed.), *The crosslinguistic study of language acquisition, Volume 1: The Data* (pp. 687-783). New Jersey: Lawrence Erlbaum Associates, Publishers.

Cohen, J.D., MacWhinney, B., Flatt, M., and Provost, J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. Behavioral Research Methods, Instruments, and Computers, *25*, 257-271.

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review, 12,* 769-786.

Corbett, G. (1991). *Gender*. New York: Cambridge University Press.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning & Verbal Behavior, 19,* 450-466.

De Bot, K. (1992). A bilingual production model: Levelt's 'speaking' model adapted. *Applied Linguistics*, *13*, 1-24.

Dewaele, J. M., & Véronique, D. (2000). Relating gender errors to morphosyntax and lexicon in advanced French interlanguage. *Studia Linguistica, 54*, 212-224.

Dewaele, J. M., & Véronique, D. (2001). Gender assignment and gender agreement in advanced French interlanguage: a cross-sectional study. *Bilingualism: Language and Cognition*, *4*, 275-297.

Donaldson, B. C. (1981). *Dutch Reference Grammar*. The Netherlands: Martinus Nijhoff.

Engle, R. W., Cantor, J., & Carullo, J. J. (1992). Individual differences in working memory and comprehension: A test of four hypotheses. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 972-992.

Franceschina, F. (2005). Fossilized Second Language Grammars. Philadelphia: John Benjamins Publishing Company.

Gougenheim, G. (1958). *Dictionnaire fondamental de la langue française.* Paris: Didier.

Guillelmon, D., & Grosjean, F. (2001). The gender marking effect in spoken word recognition: The case of bilinguals. *Memory and Cognition, 29*, 503-511.

Hagoort, P., & Brown, C. M., & Osterhout, L. (1999). The neurocognition of syntactic processing. In C. M. Brown & Hagoort (Eds.), *The neurocognition of language* (pp. 273-316). Oxford: Oxford University Press.

Hagoort, P. & Brown, C. M. (1999). Gender Electrified: ERP evidence on the syntactic nature of gender processing. *Journal of Psycholinguistic Research, 28,* 715-728.

Hardison, D. (1992). Acquisition of grammatical gender in French: L2 learner accuracy and strategies. *The Canadian Modern Language Review, 48*, 292-306.

Harley, B. (1979). French gender 'rules' in the speech of English-dominant, French-dominant and Monolingual French-speaking children. *Working Papers in Bilingualism, 19,* p. 129-156.

Hohlfeld, A. (2006). Accessing grammatical gender in German: The impact of gender-marking regularities. *Applied Psycholinguistics, 27,* 127-142.

Holmes, V. M., & Dejean de la Bâtie, B. (1999). Assignment of grammatical gender by native speakers and foreign learners of French. *Applied Psycholinguistics, 20*, 479-506.

Holmes, V. M., & Segui, J. (2006). Assigning grammatical gender during word production. *Journal of Psycholinguistic Research, 35*, 5-30.

Hyltenstam, K., & Abrahamsson, N. (2003). Maturational constraints in SLA. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 539-588). Malden, MA: Blackwell Publishing.

Jiang, N. (2000). Lexical representation and development in a second language. *Applied Linguistics, 21,* 47-77.

Karmiloff-Smith, A. (1979). *A functional approach to child language.* Cambridge, UK: Cambridge University Press.

Kroll, J. F., & Stewart, E. (1994). Category inferences in translation and picture naming: evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language, 33,* 149-174.

La Heij, W., Mak, P., Sander, J., & Willeboordse, E. (1998). The gender-congruency effect in picture-word tasks. *Psychological Research, 61*, 209-219.

Lapkin, S. & Swain, M. (1977). The use of English and French cloze tests in a bilingual education program evaluation: validity and error analysis. *Language Learning, 27,* 279-314.

Larsen-Freeman, D., & Long, M. H. (1991). *An introduction to second language acquisition research.* New York: Longman Inc.

Lenth, R. V. (2006). Java Applets for Power and Sample Size [Computer software]. Retrieved October, 1, 2008, from http://www.stat.uiowa.edu/~rlenth/Power

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22,* 1-75.

Locker, L., Hoffman, L, & Bovaird, J. A. (2007). On the use of multilevel modeling as an alternative to items analysis in psycholinguistic research. *Behavior Research Methods, 39*, 723-730.

Lyster, R. (2006). Predictability in French gender attribution: A corpus analysis. *Journal of French Language Studies, 16,* 69-92.

Marinova-Todd, S. (1994). *Critical period in second-language learning: the case of gender.* Unpublished honours thesis, York University, Toronto, Ontario, Canada.

Marx, E. (1999). Gender processing in speech production: Evidence from German speech errors. *Journal of Psycholinguistic Research, 28,* 601-621.

Miozzo, M., & Caramazza, A. (1999). The selection of determiners in noun phrase production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 907-922.

Miozzo, M., Costa, A., Caramazza, A. (2002). The absence of a gender congruency effect in Romance Languages: A matter of stimulus onset asynchrony? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 388-391.

New B., Pallier C., Ferrand L., & Matos, R. (2001). Une base de données lexicales du français contemporain sur internet: Lexique 3.02, *L'Année Psychologique, 101*, 447-462. http://www.lexique.org

O'Grady, W., de Guzman, V, M., Aronoff, M., & Rees-Miller, J. (2001). Morphology: The analysis of word structure. In W. O'Grady & J. Archibald (Eds.), *Contemporary linguistics: An introduction* (pp. 229-181). Boston: Bedford/St. Martin's.

Oliphant, K. (1998). Acquisition of grammatical gender in Italian as a foreign language. *The Canadian Modern Language Review, 52*, 239-262.

Pienemann, M. (1998a). *Language Processing and second language development: Processability Theory.* Philadelphia: John Benjamins Publishing Company.

Pienemann, M. (1998b). Developmental dynamics in L1 and L2 acquisition: Processability Theory and generative entrenchment. *Bilingualism: Language and Cognition, 1*, 1-20

Riegel, M., Pellat, J. C., & Rioul, R. (1994). *Grammaire méthodique du français*. Paris: Presses Universitaires de France.

Rummer, R. (2004). Immediate and delayed recall of visually presented sentences: Evidenc for the involvement of phonological information. *Experimental Pyschology, 51,* 15-23.

Sabourin, L. & Stowe, L. (2004). Memory effects in syntactic ERP tasks. *Brain and Cognition, 55,* 392-395.

Sabourin, L. & Stowe, L. A. (2008). Second language processing: When are first and second languages processed similarly? *Second Language Research, 24,* 397-430.

Sabourin, L., Stowe, L. A., de Haan, G. J. (2006). Transfer effects in learning a second language grammatical gender system. *Second Language Research, 22,* 1-29.

Schriefers, H. (1993). Syntactic processes in the production of noun phrases. *Journal of Experimental Psychology, 19*, 841-850.

Schriefers, H., & Jescheniak, J. D. (1999). Representation and Processing of Grammatical Gender in Language Production: A review. *Journal of Psycholinguistic Research, 28*, 575-600.

Schriefers, H. Jescheniak, J. D., & Hantsch, A. (2002). Determiner selection in noun phrase production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 941-950.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory, 6,* 174-215.

Surridge, M. E. (1989). Le facteur sémantique dans l'attribution du genre aux inanimés en français. *Canadian Journal of Linguistics/Revue Canadienne de Linguistique, 34,* 19-44.

Surridge, M. E. (1993). Gender assignment in French: the hierarchy of rules and the chronology of acquisition. *IRAL, 31*, 77-95.

Surridge, M. E. (1995). *Le ou La? The Gender of French Nouns.* Philadelphia: Multilingual Matters Ltd.

Teschner, R. V., & Russell, W. M. (1984). The gender patterns of Spanish nouns: An inverse dictionary-based analysis. *Hispanic Linguistics, 1,* 115-132.

Tokowicz, N. & MacWhinney, B. (2005. Implicit and explicit measures of sensitivity to violations in second language grammar: An event-related potential investigation. *Studies in Second Language Acquisition, 27,* 173-204

Tucker, G. R., Lambert, W. E., & Rigault, A. A. (1977). *The French speaker's skill with grammatical gender: An example of rule-governed behavior.* Paris: Mouton.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language, 28,* 127-154.

Valdman, A. (1976). *Introduction to French Phonology and Morphology*. Rowley, MA: Newbury House Publishers, Inc.

van Turennout, M., Hagoort, P., & Brown, C. (1999). The time course of grammatical and phonological processing during speaking: evidence from event-related brain potentials. *Journal of Psycholinguistic Research, 28*, 649-676.

Vigliocco, G., Antonini, T., & Garett, M. F. (1997). Grammatical gender is on the tip of Italian tongues. *Psychological Science, 8*, 314-317.