The Pennsylvania State University

The Graduate School

College of the Liberal Arts

MORPHOLOGY IN BILINGUAL LANGUAGE PROCESSING:

THE ROLE OF SECOND LANGUAGE PROFICIENCY IN ACQUIRING GRAMMATICAL GENDER

A Dissertation in

Psychology

by

Susan Christine Bobb

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

December 2008

The dissertation of Susan Christine Bobb was reviewed and approved* by the following:

Judith F. Kroll
Distinguished Professor of Psychology, Linguistics, and Women's Studies
Dissertation Advisor
Co-Chair of Committee

Carrie N. Jackson
Assistant Professor of German and Linguistics
Co-Chair of Committee

Daniel J. Weiss
Assistant Professor of Psychology and Linguistics

Maya Misra
Assistant Professor, Communication Science and Disorders

Paola E. Dussias
Associate Professor of Spanish, Linguistics and Psychology

Melvin M. Mark
Professor of Psychology
Head of the Department of Psychology

*Signatures are on file in the Graduate School

ABSTRACT

The goal of the present study was to further clarify constraints to language learning and help address questions about L2 learning that have not yet been fully resolved. The experiments examined the degree to which L2 learners and proficient bilinguals are able to fully access grammatical and morphological features of the L2. The specific aim of the study was to identify the ability of intermediate and advanced English-German bilinguals to comprehend the assignment of grammatical gender and to interpret the meaning of compounds. Grammatical gender is a feature that is typically considered difficult to acquire in the L2. Particularly for those whose native language does not mark gender, such as English, the question has been raised whether full acquisition of gender can take place and under which circumstances.

Experiment 1 set out to investigate the sensitivity of English-German and German-English L2 learners to grammatical gender and introduced the paradigm of translation recognition with simple nouns as a way to investigate gender processing. Results indicated that English-German participants had particular difficulties in rejecting correct noun translations with the wrong gender, and proficiency did not modulate these effects. In contrast, German-English participants showed robust gender effects, in which participants took longer to reject wrong translations whose gender matched the gender of the correct translation compared to translations whose gender did not match that of the correct translation. Results suggest that native speakers of German are sensitive to gender matches and mismatches across translations, and leave open the possibility that L2 learners of German who achieve native-like language competency may eventually begin to show sensitivity to gender using this task. Data from event-related potentials with

English-German participants corroborated these findings, showing no statistical support for sensitivity to gender in noun processing, and underscoring the sensitivity of L2 learners of German to semantics in translation. Data from a metalinguistic gender assignment task, however, suggested that both English-German L2 learners and German-English L2 learners were sensitive to the phonological gender distribution in German, and L2 learners of German may use these distributions as a way to behaviorally approximate native-like gender use. In a final step, morphological processing in compounding was investigated, and results for both language groups revealed sensitivity in processing internal gender agreement in compounds, although the pattern of data were not in the predicted direction. Together, the results of these experiments confirm previous results on the difficulty of L2 gender processing in German (e.g., e.g., Sabourin, Stowe, & de Haan, 2006) and also appear to show dissociations between tasks that require more automatic processing and those that are under the participant's control.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

Many thank yous are in order, and this page will not begin to sufficiently express my gratitude and indebtedness. I would first like to acknowledge funding by a National Science Foundation Dissertation Research Grant BCS-0720155, a teaching release dissertation award from the College of the Liberal Arts Research and Graduate Studies Office at the Pennsylvania State University, a German Academic Exchange Program (DAAD) short term research grant, and the Eileen Wirtshafter and Herschel W. Leibowitz Graduate Travel Scholarship from the Department of Psychology at the Pennsylvania State University.

I am very grateful for having an exceptional doctoral committee and wish to thank Judy Kroll, Carrie Jackson, Giuli Dussias, Maya Misra, and Dan Weiss for their wisdom, availability, and patience throughout this dissertating process. Judy, I feel deeply honored to have learned from you the past five years as my advisor and dissertation chair – it has been a true privilege. Not only are you an incredibly gifted researcher, you are an advocate for those around you, and have a generosity in spirit that I aspire to. Thank you for your guidance, for seeing possibilities when I could not, for investing so much in me. Carrie, thank you for co-chairing my dissertation committee. The research is the richer for your involvement, and I have learned a lot about doing research and about writing through your wonderful mentorship.

To the friends and colleagues whom I met along the way:

Anat Prior, Tamar Degani, Zofia Wodniecka, Philipp Kanske, Kathrin Rothermich, Moritz Dannhauer, Eva Suarez Büdenbender, Matt Carlson, Rose Halterman, Tracy Kramer Scaltz, Emily Rine, Jonathan Rhinehart, Stefanie Rehn Jordan, and especially Paula Golombek, who offered invaluable insights, costly time and resources, as well as their generous friendship.

Maya Misra, Taomei Guo, Sonja Kotz, Sylvia Stasch, Dorothee Geier, who first introduced me to the world of ERPs and taught me its methodology.

Thom Gunter, Janet van Hell, Rosa Sanchez-Casas, Dirk Koester, Laurie Beth Feldman, Wouter Duyck, and many more for helpful resources and discussions in designing the dissertation study.

Kate Chang, Tyler Phelps, Jared Linck, April Jacobs, Cari Bogulski, Rhonda McLane, my lab mates, past and present. A special thank you to my "big sister," long time officemate, and dearest friend Noriko Hoshino for mentoring me and walking along side me the past five years.

To my past advisors, Tim Loboschefski at Sweet Briar College for sparking my interest in psychology, Marie-Thérèse Killiam at Sweet Briar for encouraging my love of literature and languages, and John Bonvillian at the University of Virginia for showing me that I could combine my love of languages with my fascination with the workings of the mind.

To my research assistants Angélique Blackburn, Mark Connelly, Shara Gress, and Hannah Morton, as well as Jason Gullifer, Mark Minnick, and Joyce Tam, for their help in running participants, cleaning up data, and organizing the lab. This project would not have happened without you in the trenches with me.

To the incredible staff in the Department of Psychology: Amy Larimer, Connie Moore, Deb Auman, and Jean Courter, who do what they do so well, so that I could focus on my research.

To my off-site hosts: Natasha Tokowicz at the University of Pittsburgh, and Angela Friederici at the Max Planck Institute for Cognition and Neuroscience, Leipzig, Germany for supplementing the small pool of bilinguals here at Penn State. Natasha, thank you, too, for all of your advice in statistics and ERP methods.

Thank you to my family, who not only supported me from a distance, but jumped right into my research, participating as my pilot subjects, as well as helping to edit some of the first drafts of this dissertation. Dad, Mom, Steph, Bart, Amanda, Audrey, and Daphne, I love you dearly. Thank you for inspiring me to pursue this dream, believing for me that it could be attained in the moments when I was ready to give up. Grandma, thank you for our encouraging phone conversations, especially during the last few months of writing. Thank you also to my new family, Dave and LeeAnne, Travis, and Phil, for their warm welcome and continuous encouragement the past year and a half.

To my State College community at large, both past and present: The Laytons, the Bowmans, the Hesters, the Hesses, the Letourneaus, Griffin and Sharon Jones, Megan Colabrese , Ellen (Key) and Ed Ballock, Chris Anderson, Morgan Forney, Caitlin Forney, Maria Morck, Mary Kay Horton, LaNette Dellinger, and so many more. And Betsy (Stiver) Forney, one of the most incredible roommates I have had the privilege to live with. I owe so much of who I have become to each of you and see God so clearly through your lives.

To my now permanent roommate, the love of my life, my husband Jamie Clark, I would not have crossed the finish line without you by my side. I'm looking forward to many more adventures together!

Solo Deo Gloria.

Chapter 1

Introduction

The adaptivity of the brain is such that we spend our lifetime learning, reorganizing neural pathways in response to new experiences in ways that allow us to acquire new knowledge and new skills well into advanced age. From a child's first word and first step, to a college student's grasp of statistics and a grandmother's first email message, our brains are wired to continually learn. This neural plasticity is showcased nowhere better than in the acquisition of a foreign language, particularly after the onset of puberty when aging presumably limits ultimate language proficiency. Nonetheless, and in spite of these constraints, adults *can* learn a second language, although anyone having attempted, let alone succeeded, in such a feat will readily recount the numerous obstacles they had to overcome in the process. That any such learning can take place at all is startling given the complexity of the task: Learning a foreign language requires more than mapping new word forms onto previous concepts and slotting them into the native language grammatical system. In most cases, learning a foreign language to even a rudimentary level requires understanding new grammatical structures, not to mention phonological and pragmatic considerations. And still, native speakers of languages as diverse as Mandarin, French, Xhosa, Tamil, Spanish, Arabic, German, Urdu, and English are able to learn each other's languages.

The past two decades has seen a marked rise in research focusing on bilingualism and second language learning and the underlying cognitive processes that enable even adult language learners to become skilled users of a given language. Several studies have investigated the possibility of language transfer from a first language (L1) to a second language (L2). Fewer have systematically investigated the acquisition of structures that are particularly difficult for non-native speakers to acquire and that do not exist in the native language, as is the case for grammatical gender in English. Using converging evidence from behavioral and event-related potential (ERP) measures, the goal of the present thesis is to examine whether the apparent insensitivity to a grammatical feature of the L2, such as gender, reflects a fundamental representational absence of that feature or rather an inability to utilize that grammatical information in L2 language processing.

While this thesis focuses on L2 acquisition of grammatical gender, I do not want to lose sight of the broader picture, which is to explore the extent and limits of L2 learning. Rather than an end in itself, investigating these structures becomes a window into the acquisition process as a whole in the hopes of providing clues into the dynamics of late L2 learning.

Chapter 2 will focus on issues of plasticity and constraints in second language acquisition (SLA), including age of acquisition effects and potential L2 learning mechanisms. I will also address the representation and processing of grammatical gender in both first and second language. In the following chapter, Chapter 3, the general motivations and predictions of the current studies will be presented, including an overview of the methods and paradigms that were used. The experiments investigating gender processing in simple nouns will be presented first in Chapters 4 and 5, and the behavioral paradigm will be extended to ERPs in Chapter 6. I will then address theoretical and experimental issues in compounding and gender processing in Chapter 7, before presenting the results of a compounding experiment. The final chapter then attempts to integrate present findings with past research to provide potential implications for gender processing, in particular, and second language processing more generally.

I would like to add a final note about terminology. In the following thesis, the term “gender” will be used to refer to grammatical gender, and when the natural gender inherent to a person or animate object is intended, I will use the term “biological gender”. In keeping with current conventions in psycholinguistic literature, the term “bilingual” will be used to describe any proficient second language user while “language learner” will indicate any speaker on a broad spectrum of the learning continuum.

Chapter 2

Plasticity and Constraints in Second Language Learning

In countries like the United States, exposure to a second language during childhood is becoming increasingly prevalent; however, most students do not seriously undertake the study of a foreign language until high school or college[1]. Anecdotal and research evidence both suggest that the later in life learners are exposed to the second language, the less successful the learners are in attaining the L2 grammar, and especially the phonology. Accented speech is a common trait in the late L2 learner, and particular grammatical features seem to show persistent difficulties well after the learner has achieved so-called conversational proficiency. In the present study in which participants are late L2 learners (i.e., first contact with the L2 after the onset of puberty), age is a particularly important consideration in characterizing patterns of language use in these adult learners. Although early research on age of acquisition (AoA) paid little attention to specific language structures, more recent research has begun to show it is precisely by investigating these specific structures that constraints in late L2 learning emerge. The following literature review will detail the emergence of these constraints, focusing on one grammatical feature, namely gender, that has proven to be one of the most difficult features for late L2 learners to learn.

[1] A notable and increasingly prevalent exception which I will later touch on is heritage speakers or children of immigrants.

2.1 Age of Acquisition Effects

When considering constraints on second language learning, there is the distinct possibility that the very age at which language learning begins places specific limitations on the ultimate proficiency in that language. The idea of a critical period, of a bounded time span within which language may be fully acquired, has fueled research to inform the debate on factors determining the plasticity of the language system over one's life span. To date, however, research findings remain mixed and even suggest that some aspects of language, and even some specific language structures, seem more prone to AoA effects than others (for more extensive reviews of the critical period hypothesis, see Birdsong, 2005, and DeKeyser & Larson-Hall, 2005; see also Hyltenstam & Abrahamsson, 2001; Marinova-Todd, Marshall, & Snow 2000, 2001, as well as DeKeyser, 2000, and Bialystok, 2002).

Lenneberg (1967) first implicated age as a determinant in successful language learning in his book, *Biological Foundations of Language* (see also Penfield & Roberts, 1959). In his work, Lenneberg addresses, among other topics, the concept of a critical period for language, with an emphasis on the proposal that there are biological underpinnings reflected in maturational development which determine this process. Where a lack of development limits early learning, the decline in brain plasticity hinders learning at later ages (cf. p. 179). Other research has critized Lenneberg's proposal, however, because it is vague with respect to the specifics on how this critical period limits second language learning (e.g., Johnson & Newport, 1989; Harley & Wang, 1997):

Lennenberg notes that “language-learning-blocks” severely hinder successful second language learning into adulthood, but leaves their exact nature underspecified.

Johnson and Newport (1989) were among the first to empirically test the idea of a critical period in second language learning. In their now seminal work, Johnson and Newport studied Korean and Chinese immigrants to the United States, testing them on a range of language knowledge from English morphology to syntax using a sentence grammaticality judgment task. In the study, age of arrival predicted the immigrants’ level of English proficiency, with arrival before the age of 6 predicting native-like command of English. Arrival between the ages of 7 and 15, however, correlated negatively with proficiency, and arrival after 15 showed no relationship between age and proficiency. The authors interpreted the results as suggesting a critical period for L2 learning that ended by the age of 15.

Other studies, however, raise doubt about the exact upper limit of a critical period for L2. In a reanalysis of Johnson and Newport’s data, Bialystok and Hakuta (1994) found that a cut-off age of arrival of 20 characterized the boundary of learning equally well as Johnson and Newport’s original boundary of 15. Birdsong and Molis (2001) reported a continuing decrease in L2 learning with increasing age, rather than a specific age cut-off, and also showed evidence for L1-L2 pairing effects. These two findings, a cut-off age which is outside of puberty and even the finding of a lack of an exact cut-off age, challenge the traditional view of the critical period hypothesis by which learning is maturationally bounded by puberty.

In all three of these studies, the increase in variability among the late age of arrival group compared to the early arrival group suggests the presence of other variables in determining successful L2 learning. While age continues to be the best predictor in these studies, Birdsong and Molis (2001) make the case that other factors should be further investigated, such as how much the target language is used and what types of structures are tested. In this last respect, the results of Flege, Yeni-Komshian, and Liu (1999) are particularly relevant to the present study, showing that age of arrival correlated most with grammatical proficiency in the use of irregular morphology, not regular morphology. Flege et al. tested Korean learners of English on a subset of items from the original Johnson and Newport (1989) study and divided the items into rule-based and lexical items. While participants continued to be highly accurate in judging rule-based items regardless of age of arrival, lexical items showed a marked decrease in accuracy as a function of arrival age. The results suggest that those aspects of the grammar which are less systematic were those most susceptible to age effects (see also Birdsong & Flege, 2001).

Similarly, as seen throughout various studies, phonology seems to be more affected by AoA than morpho-syntax (e.g., Flege, Yeni-Komshian, & Liu, 1999; Sebastián-Gallés, Echeverría, & Bosch, 2005). Perhaps most interesting of all is the finding of near-native L2 proficiency among some of the second language learners (e.g., Birdsong, 1992; Birdsong & Molis, 2001; Bongaerts, 1999). While the findings of "near-nativeness" must in some ways be qualified by how the authors defined "near-nativeness" in these studies, the existence of high proficiency among late learners suggests at a basic level that in some cases learners can overcome the consequences of age.

The evidence thus suggests that the strong version of a critical period in L2 learning, in which there is a non-linear relationship between age and learning and where native-like learning of the L2 is not possible after puberty may not exist. Rather, a linear relationship between age and learning continues beyond puberty, indicating the continuous influence of factors such as cognitive decline, psychosocial factors, as well as a loss of brain plasticity in the attainment of the L2. The current study, while focusing on one particular grammatical feature, will attempt to address some of these constraints by obtaining measures of cognitive resources in addition to a detailed history of language learning. In this manner, I hope to be able to begin to tease apart some of the independent contributions to late L2 learning made by individual differences in working memory and cognitive control.

2.1.1 Evidence on Age Effects

While controversy still exists over whether there is a critical period per se in L2 learning, it does appear that age can be one factor influencing the rate and extent to which a second language is learned. Two general lines of research have attempted to further address the consequences of age and proficiency in learning a second language, increasingly trying to disentangle their effects on specific areas of language (i.e., syntax, morphology, phonology). Neuroscientific studies, the first to be reviewed here, have examined brain activity using tools such as functional magnetic resonance imaging (fMRI), positron emission tomography (PET), and event-related potentials (ERPs) in order to gain a better understanding of how L1 and L2 activation converge or diverge in the brain. Behavioral studies, the second type of studies to be reviewed, use speed of response (reaction time in milliseconds) and response accuracy in experimental tasks as a window into cognitive processing. The following section presents relevant research from both types of studies as it pertains to constraints and plasticity in L2 learning.

2.1.2 Neuroscience Studies

Neuroscientific evidence suggests extreme examples in both plasticity and constraints to L2 learning. Methods used in these approaches show sensitivity to electrical potentials (event related potentials or ERPs), blood oxygen levels (functional magnetic resonance imaging or fMRI), or glucose metabolism and regional cerebral blood flow (positron emission tomography or PET). Researchers can capitalize on these sensitivities to ask the question of how the brain processes a second language and better understand underlying brain mechanisms.

Due to its high temporal resolution, the recording of ERPs has provided important evidence on the neural basis of language processing. Two early ERP studies, Weber-Fox and Neville (1996) and Hahne and Friederici (2001), support the general claims of a critical period, illustrating the constraints that late L2 learning places on ultimate attainment of the L2. In ERPs, electrical brain activity is measured from the scalp and time-locked to the onset of stimulus presentation. By averaging trials of stimuli with similar linguistic characteristics, noise from other sources is reduced, producing a waveform reflecting the processing of the linguistic characteristic of interest. Weber-Fox and Neville specifically investigated sentences with semantic and syntactic violations such as:

(a) The scientist criticized Max's *event* of the theorem. (semantic-violation)
(b) The scientist criticized Max's *of* proof the theorem. (syntactic-violation)

(c) The scientist criticized Max's proof of the theorem. (control sentence)

Typically, native-speakers of English reading sentences such as (a) show an N400, a negative-going waveform around 400 ms after stimulus onset, which in this case is the italicized word *event*, as compared to a baseline control sentence, such as to "proof" in (c). In contrast, sentences such as (b) typically show a P600, a positive-going waveform around 600 ms after stimulus onset. Chinese-English participants in Weber-Fox and Neville's study ranged in their initial age of L2 exposure from 1-3, 4-6, 7-10, 11-13, and after 16 years of age. Results of the experiment revealed robust age effects in the ERPs for syntactic violations, such as (b), in participants as young as 1 to 3 years of age. In contrast, semantic differences between the L2 learners and the English native speaker controls only started to appear in participants arriving after the age of 11, so that before the age of 11, semantic processing of the bilinguals looked similar to that of native-English speakers. The results thus suggest that the age of L2 learning affects syntactic processing more than semantic processing.

In another ERP study, Hahne and Friederici (2001) showed a similar divergence between semantic and syntactic processing in an auditory listening task for native Japanese bilinguals who were late L2 learners of German. Consistent with Weber-Fox and Neville (1996), there was an N400 effect for semantic violations similar in form and distribution to that of native German speakers, while the P600 for syntactic violations did not show significant modulations, suggesting non-native-like syntactic processing.

While the studies by Weber-Fox and Neville (1996) and Hahne and Friederici (2001) provide evidence that supports the general critical period claims, other research suggests no lasting effects of age on L2 learninig, and even limited evidence that the L2 can completely replace the L1. Pallier et al. (2003) investigated language development of Korean adoptees by French parents in France, thus taking advantage of an extreme situation in which learners were completely cut off from their L1 early in life. If AoA effects reflect true maturational constraints, then exposure during the first few years of life should be critical, showing both retention of the earlier learned Korean, as well as less than native-like processing of French. The adoptees, whose ages ranged between 20 and 32 at the time of testing, were adopted between the ages of 3 to 8 years, and all reported having forgotten their maternal language, Korean. fMRI results showed no differential brain activation for Korean compared to other unknown foreign languages, suggesting that their initial L1 had been completely replaced by their L2, French. Compared to native French speakers, the adoptees showed similar, although more restricted, activation patterns when exposed to French language stimuli. While this study seems to hold more direct implications for "L1 replacement" as opposed to "L2 addition", these results, as revealed by similar brain activation patterns between the monolingual/L2 French participants and the Korean adoptees, do seem to suggest that a second language can reach native-like status, at least when the L1 is not maintained. Other research similarly shows that early bilinguals use identical brain areas for language processing in L1 and L2 (e.g., Hernandez, Martinez, & Kohnert, 2000; Klein, Milner, Zatorre, Meyer, & Evans, 1995; but see also Perani al., 2003, suggesting subtle differences). Other evidence, however, challenges the results of Pallier et al., showing enduring benefits in pronunciation later in life as a result of early language exposure (Au, Knightly, Jun, & Oh, 2002; Knightly, Jun, Oh, & Au, 2003; Oh, Jun, Knightly & Au, 2003; Venturevra, Pallier, & Yoo, 2004).

Studies, such as Perani et al.'s (1998) PET study, question whether results such as those by Weber-Fox and Neville (1996) and Hahne and Friederici (2001) depend on the age of learning. The results of Perani et al. show similar evidence of brain plasticity for late bilinguals where highly proficient late bilinguals (Italian-English) activate brain regions identical to highly proficient early bilinguals (Spanish-Catalan). Late bilinguals with lower proficiency, on the other hand, do not (cf. Perani et al., 1996). Importantly, this finding was for a listening-comprehension task, and may suggest that the type of language task may show differential effects (see Kim, Relkin, Lee, & Hirsch, 1997, for the opposite pattern in a sentence generation task).

Other neuroimaging studies have produced a mixed pattern of results, providing evidence for both constraints and brain plasticity, often within the same study. Of particular interest is a study by Wartenburger, Heekeren, Abutalebi, Cappa, Villringer, and Perani (2003), which shows different patterns of activation for grammatical and semantic processing. In this fMRI study, three groups of Italian-German bilinguals (early learners with high proficiency, late learners with high proficiency, and late learners with low proficiency) were tested on grammaticality and semantic judgment tasks. Only early bilinguals activated similar grammatical areas to native German speakers, while highly proficient bilinguals activated similar semantic areas to native German speakers, regardless of L2 AoA. These results suggest differential effects of proficiency and AoA on L2 grammar and semantic development similar to Weber-Fox and Neville (1996), highlighting the fact that grammatical knowledge seems to be particularly difficult to acquire later in life, a point which will be returned to later in this thesis.

The imaging studies reviewed here and elsewhere (e.g., Abutalebi & Green, 2007) show extended brain activation in L2 processing specifically for areas implicated in cognitive control, suggesting increased processing difficulties compared to native speakers. At the same time, images from grammatical and semantic judgment tasks also show activation differences between native and L2 language speakers which would implicate qualitatively different, in addition to more effortful, L2 processing (e.g., Abutalebi & Green, 2007; Wartenburger et al. 2003; Rueschemeyer, Fiebach, Kempe, & Friederici, 2005).

Other studies investigating sentence processing using electrophysiological measurements also support these conclusions. In a recent study using event-related potentials (ERPs), Rossi, Gugler, Friederici, and Hahne (2006) investigated the processing of morphosyntactic agreement and word category violations in late L2 learners of German and late L2 learners of Italian with either high or low proficiency. One of the components which they were expecting was a left anterior negativity (LAN) which is implicated in morphosyntactic processing such as morphosyntactic violations, as well as a P600, reflecting integration and reanalysis of syntactic information. Results suggested that while low proficiency bilinguals showed qualitatively and quantitatively different processing (an absence of a LAN, and a delayed P600 respectively), high proficiency bilinguals of both languages showed equivalent processing to native speakers. There were no temporal or amplitude differences between the two groups. These results contrast with earlier results showing constraints to L2 syntactic processing

(Hahne, 2001; Hahne & Friederici, 2001; Weber-Fox & Neville, 1996). One possible reason for the differences in results could be due to the nature of the materials. While previous studies had used passive sentences, the study by Rossi et al. presented sentences in the active voice, which tend to be less complex than passive constructions. If the type of material is truly the source of the different patterns, the results provide more evidence for processing load, and not a critical period, in limiting native-like L2 processing.

Recent research has also qualified the initial support for native-like semantic processing. While previous studies specifically investigated sentence structures with semantic anomalies, a more recent study that did not focus on such a "violation paradigm" has found more nuanced results in the extent to which L2 semantic processing can become native-like. Kotz and Elston-Güttler (2004) specifically addressed this concern in a priming paradigm using a lexical decision task (LDT) with late German-English bilinguals. Typical priming is tested by presenting word pairs such as boy – girl and asking participants to judge whether the pair is related or not. In contrast to semantic anomaly testing, LDT presents lists of words to which participants must decide whether they are words or nonwords. Within these lists, primes and targets are imbedded, allowing the testing of priming without presenting overt word pairs as used in other research paradigms. As a result, the LDT task provides a measure for testing more automatic processing that is less influenced by participant-controlled strategy effects (cf. Shelton & Martin, 1992).

Kotz and Elston-Güttler (2004) investigated two types of semantic priming: associative and categorical. In associative priming, presentation of a prime word that is associatively related to the target (i.e., a relationship between terms in which one term leads to the other) speeds decision making on the subsequent target word (e.g., boy – girl). Categorical priming (i.e., a relationship between terms in which both terms belong to the same category but do not lead to one another) shows speeded decisions on the target when the prime is categorically associated with the target (e.g., boy – junior). Results indicated that advanced and lower proficiency bilinguals indeed showed differential processing. Both groups showed associative priming in their ERP results; however, advanced bilinguals displayed greater sensitivity to categorical priming than lower proficiency bilinguals. But even the categorical priming in the advanced group did not approximate that of native English speakers (see Ardal, Donald, Meuter, Muldrew, & Luce, 1990, for other fine-grained differences in semantic processing for the L2, as well as Weber-Fox & Neville, 1996, for AoA effects). Interestingly, these ERP results contrast with behavioral results from the same task, in which the advanced group showed categorical priming similar to that of native speakers.

The evidence at the neural level thus seemed mixed, showing both constraints and plasticity as a function of proficiency and the age at which the L2 was first learned. While syntactic processing appears particularly vulnerable to age and proficiency effects, even semantic processing shows constraints. Once thought more open to native-like processing, L2 semantic processing, upon closer examination, also differs from L1 processing. These results suggest that while L2 learners may achieve native-like behavior, they may never achieve native-like processing, maintaining instead persistent fine-grained differences between the L1 and the L2.

2.1.3 Behavioral Evidence

While neuroscientific research supports the advantages of early L2 learning, some behavioral evidence suggests that even early L2 learning will produce differential outcomes compared to L1 learning. Earlier research by Pallier and colleagues on speech perception suggests that even slight delays in early L2 language learning can affect the L2, distinguishing the L2 speaker from native speakers of the same language (Pallier & Sebastiàn-Gallés, 1997; see also Bosh & Sebastiàn-Gallés, 1997; Bosh & Sebastiàn-Gallés, 2003; Sebastiàn-Gallés & Soto-Faraco, 1999). In the Pallier and Sebastiàn-Gallés (1997) study, Spanish-Catalan bilinguals with exposure to Catalan by the age of six performed an acoustical classification task in which they had to decide whether words contained [e] or [ɛ], two sounds which are phonemically distinct in Catalan, but not in Spanish. The results of the classification task and a following discrimination task showed sensitivity among Catalan monolinguals, but not Spanish-Catalan bilinguals, in discriminating between these sounds. Participants then judged the previous stimuli as to their categorical fit with Spanish [e], Catalan [e], or Catalan [ɛ]. Results of the typicality judgment task continued to reveal differences in vowel processing between the two groups.

In contrast, research by Minagawa-Kawai, Mori, and Sato (2005) with early Korean-Japanese bilinguals showed no behavioral differences in categorical perception between bilinguals and native speakers in discriminating Japanese long/short vowel distinctions. Near-infrared spectroscopy, however, revealed differential cortical processing of Japanese for each group.

While fine-grained differences exist even for early L2 learners, at least for phonology, there is also evidence that early L2 learning, even if discontinued for a while, conveys benefits for later L2 phonological and grammatical development. A recent study by Au and colleagues (Au, Oh, Knightly, Jun, & Romo, 2008) specifically investigated childhood learners of Spanish who then did not use the language until later in puberty (around 14 years of age). Compared to childhood overhearers (i.e., children who overhear a language growing up, but do not actively use the language), these language learners not only showed more native-like phonology, but also more native-like grammar, as measured by noun- and verb-phrase production and a sentence grammaticality-judgment task. These results seem to contrast with the earlier results of Pallier et al. (2003) in which Korean adoptees showed no active memory of their L1, Korean. However, in the Au et al. (2008) study, the childhood speakers still had some language support for their L1 in that participants were childhood speakers of Spanish living in Southern California, who still had contact with Spanish-speaking relatives, although they themselves rarely if ever spoke Spanish between the ages of 7 and 12. In addition, at the time of testing, they had already gone through a phase of "relearning", taking Spanish classes in high school. In contrast, participants in the Pallier et al. study had neither language support in their community, nor a re-learning phase at testing (for results supporting the idea that childhood speakers have an advantage over novice Korean learners in re-learning see Oh, Jun, Knightly, & Au, 2003).

Other recent behavioral evidence also suggests that, in many ways, L1 and L2 speakers process a given language differently, and that adult L2 processing differs from child native language processing. The main argument of Clahsen and Felser (2006) is that adult second-language learner sentence parsing does not resemble the parsing of children whose patterns reveal shortages in working memory (WM) (cf. p. 28, and Felser & Roberts, 2007). If both children and adult L2 learners place high demands on WM resources, and this factor is the main determinant in parsing strategies, one would expect their patterns of behavior to be similar. Instead, Clahsen and Felser propose a shallow-structure hypothesis by which L2 learners simply do not encode syntactic information to as much detail as adult L1 parsers. Thus, while WM plays a role in child processing, it does not play as significant a role in adult L2 sentence processing (but see Michael & Gollan, 2005; Hoshino, Dussias, & Kroll, under revision).

The picture that emerges from these data is complex, and efforts to reconcile all of these results must inevitably take into account variables such as the language pairings under investigation, the age of learning and proficiency, as well as the type of language support and learning environment the bilingual has received. While to my knowledge no one has systematically compared all of these variables, one study which begins to address these questions is Bialystok and Miller (1999) (see also Birdsong & Molis, 2001). In this study, two learner groups (Chinese-English and Spanish-English), each with two levels of AoA (younger than 15 and older than 15), were compared to native English speakers on a spoken and written grammaticality judgment task. The study not only addressed issues of AoA and proficiency, but also looked at language pairing, since the grammatical structures under investigation in English either overlapped with the participants' L1 (Chinese or Spanish) or did not: While the future tense and present progressive are formed similarly in English and Chinese, they are formed differently in Spanish. In contrast, while both English and Spanish use a similar system of determiners and plural markings, Chinese uses neither.

Results showed significant differences in responses between the two language groups, but not between age groups within a language group. Specifically, the present progressive sentences were easier for all language groups, including native English speakers, regardless of proficiency. Only the Chinese group showed an effect of the degree of language overlap for the other structures, and they exhibited this effect regardless of age group. Bialystok and Miller (1999) account for this result by citing the overall higher L2 proficiency of the Spanish group, suggesting that structural effects are most evident at earlier stages of learning. The fact that the two language groups differed in their response to language structure and were unaffected by age did not support a critical period. Bialystok and Miller also speculate that the nature of the language pairings may have allowed the Spanish group to acquire higher levels of proficiency, given equal amounts of exposure and time.

What are the implications of these results, particularly as they pertain to language pairings, for some of the other findings on L2 plasticity and constraints? A quick evaluation, both of the studies reviewed here and other studies, suggests that language pairings play a role in the findings presented here. Particularly the results of Birdsong and Molis (2001) are relevant, as they undertook a replication of Johnson and Newport (1989) with identical materials, but instead of Korean-English and Chinese-English participants, they tested a group of Spanish-English bilinguals. Their results, in contrast to those reported by Johnson and Newport, showed no evidence of a critical period and even found modest evidence for native-like grammatical attainment in some of the older participants. A review of other language pairings suggests that particularly for pairings across language families (e.g., Chinese-English, Italian-German, Russian-German, Japanese-German) that there are increased findings of constraints to L2 processing (Johnson & Newport, 1989; Wartenburger et al., 2003; Hahne 2001; Hahne & Friederici, 2001), but that early language learning (Weber-Fox & Neville, 1996, Wartenburger et al., 2003) and/or high levels of L2 proficiency (Rossi et al., 2006; Wartenburger et al., 2003) attenuate these limits.

2.2 Accounting for the Difference: Models of L2 Learning

2.2.1 Linguistic Models of L2 Learning

Several models exist which make specific predictions about the extent to which L2 learning can take place and how close L2 language processing can come to native language processing. The Failed Functional Features Hypothesis (FFH, Hawkins & Chan, 1997; Hawkins & Franceschina, 2004) specifically takes a critical period into account, proposing that after the critical period, L2 learners cannot learn certain language features, or can only learn them with great difficulty. In contrast, Schwartz and Sprouse (1994, 1996) offer the Full Transfer/Full Access Hypothesis (FTFA) by which L2 learners first transfer their entire L1 grammar into their L2 and subsequently draw on full access to Universal Grammar (UG) to restructure parameters and acquire structures unique to the L2. While the former proposal predicts strong constraints to late L2 learning, the latter proposal suggests that native-like attainment is possible. Schwartz and Sprouse also predict that languages with more featural overlap might be more conducive to L2 learning, accelerating initial learning through transfer.

Jiang (2000) similarly suggests a transfer-based model of SLA in which L2 lexical representations depend on the transfer of L1 representations. However, only language-shared properties can be transferred so that language-specific features are not integrated into the L2 representation. Instead of being stored in the mental lexicon, semantic, syntactic, and/or morphological information may be stored in general or episodic memory, making automatic processing impossible of such features. Crucially, unlike the FTFA hypothesis, UG does not exist to allow attainment of these missing L2 specific features, forcing learners to rely on explicit linguistic knowledge in order to correctly use features such as language-specific inflectional morphology (see also Sorace, 2003).

In support of his theory, Jiang (2004, 2007) noted that Chinese L2 learners of English displayed particular insensitivity to morphological processing of the plural marker, which does not exist in Chinese, but were sensitive to other violations such as pronoun-verb agreement and verb argument structure (2007) in a moving-window reading paradigm. Off-line, however, L2 learners displayed high accuracy in a written test in using the plural marker. Jiang interprets the results in support of a Competence Deficit Approach (CDA), in which L2 learners have incomplete lexical representations in the L2 that hamper automatic, but not explicit, processing. A Performance Deficiency Approach (PDA) would have predicted difficulties in accessing or controlling already-internalized language competence. Because L2 learners showed difficulty in a receptive (i.e., reading) task, which was not under explicit control, a PDA could not account for these results (For other research arguing against native-like mental representations see Coppetiers, 1987; Liceras, 1997; Lozano, 2003; Hawkins & Liszka, 2003; for the opposite view see Birdsong, 1992; Bruhn de Garavito, 1999; Montrul & Slabakova, 2003; White & Genesee, 1996).

Finally, in his competition model, MacWhinney (1987, 1997) also focuses on transfer and interference between L1 and L2 in language learning, albeit without the influence of UG (see also Bates & MacWhinney, 1981). Instead of linguistic structure, universal cognitive structures (i.e., shared learning mechanisms) drive language learning: MacWhinney represents acquisition in a connectionist model in which learning occurs as a result of the strength and validity of different grammatical or semantic cues in the input rather than innate principles or parameters. The L2 learner initially transfers the L1 system into the L2. Over time based on the input the learner receives, the two languages then begin to diverge, with the developing L2 system becoming more native-like in its ability to process L2 input.

2.2.2 Statistical Learning Models of L2 Acquisition

What mechanisms then underlie the ability, or inability, to acquire a second language as an adult? While several proposals exist, the potential learning mechanism that I will briefly consider here is the processing system's sensitivity to statistical modes in the speech stream. In many ways statistical learning mechanisms are similar to and compatible with MacWhinney's Competition Model, and in fact could be argued to be the model's underlying learning mechanism. As MacWhinney points out, statistical regularities and co-occurrences in corpora form the basis of much of the learning in emergentist views (MacWhinney, 2006). Early linguists argued that probabilities could not play a significant role in language learning (e.g., Chomsky, 1980); however, increasing evidence in more recent years has shown that statistical learning, as opposed to or in addition to rule learning, can in fact take place and may be one of the main mechanisms underlying both first and second language learning (Ellis, 2000, 2002).

Recent evidence suggests that both child and adult language learners use statistical probabilities within and between words to extract meaningful speech segments from the speech stream. Saffran, Aslin, and Newport (1996) showed that 8-month-old infants, even after only 2 minutes of exposure to a created language, could discriminate between words and nonwords, and more importantly words and part words, by using transitional probabilities between sounds. Newport and Aslin (2004) then investigated whether adults could apply statistical learning to nonadjacent syllables or consonant/vowel segments. Interestingly, they found that participants only learned word boundaries based on consonant and vowel segments, but not based on non-adjacent syllables, a pattern that matches natural-occurring language patterns (see also Pierrehumbert, 2003, and Maye, Werker, & Gerken, 2002, for further discussion of child language learning).

Some researchers have also shown that infants are able to abstract not only words, but also grammar from speech streams (Gomez & Gerken, 2000). Data on distributional properties of language in artificial languages have shown the ability of adults to extract grammatical gender-like categories of nouns (Brooks, Braine, Catalano, Brody & Sudhalter, 1993; Frigo & McDonald, 1998) as well as syntactic properties (e.g., Billman, 1989). However, it is debated whether these patterns still comply with a statistical learning mechanism or rather show an independent algebraic rule-based learning mechanism for syntax (McClelland & Plaut, 1999; Marcus, 1999). As McClelland and Plaut point out, statistical learning mechanisms are powerful enough to generalize in such a manner so as to give the appearance of “rule-based learning.” Important for the current discussion is the finding that acquiring increasingly complex features of language, such as syntax and morphology, may depend on learning mechanisms sensitive to distributional properties of a given language and that these distributions can be generalized to account for apparent rules in a language (but see also Jackendoff, 2002; Marcus, 2001; and Pinker & Ullman, 2002, for counterarguments).

How would a statistical learning mechanism then be instantiated in bilingual or second language learning (c.f., Saffran, 2003)? One could argue that multiple sets of statistics might be confusing, or that once the statistical mechanism is tuned to and used for one language, it is unavailable when learning another language. The evidence suggests, however, that this is not the case. The work by Maye, Werker, and Gerken (2002) looking at bimodal vs. unimodal frequency distributions of speech sheds some potential light on this issue and suggests that the very distribution of speech may allow for multiple sets of statistical constraints.

Other research on adults provides evidence that statistical learning may not be a domain-specific mechanism but instead may act as part of a single segmentation mechanism that can be used for speech, as well as tone, segmentation (Saffran et al., 1999). This finding is important because, together with the evidence for statistical learning in both children and adults, it provides potential support for a more general learning mechanism unbounded by or with only limited developmental constraints into adulthood (see also Wilkins & Wakefield, 1995; Ullman, 2004, for similar perspectives regarding age constraints while still advocating general rule-based mechanisms). In this way, statistical learning appears to complement other accounts of language learning, particularly connectionist models, in late bilingual language learning. Statistical learning of novel languages in adulthood complements research suggesting successful L2 learning among late learners can proceed outside of a critical period (Saffran et al., 1999; Newport & Aslin, 2004).

The review of the literature on L2 learning as presented here paints a complex picture, suggesting both constraints as well as incredible plasticity in the learning process. Several of the models reviewed in the last two sections offer the possibility that L2 learning can proceed similarly to L1 learning. Some research, in contrast, has proposed a dissociation between automatic and non-automatic language representation (e.g., Ellis, 1984), suggesting that the outcome of L2 learning may be native-like L2 performance but not native-like processing (see also Carroll, 1989). It is possible, as Jiang (2000) points out, that reaching a more advanced stage of processing (i.e., a lexical entry with integrated semantic, syntactic, and morphological information), while difficult to attain, can occur "if sufficient, highly contextualized input in L2 is available and processed by the learner" (2000, p. 54). Some structures of language, however, seem to show persistent difficulties, regardless of the level of language proficiency, and this is the topic to which I will turn next.

2.3 Grammatical Gender in Language Processing

Grammatical gender no longer exists in English[2], so that if an English speaker were to name a cat, she would simply say that it is “a cat” or point and say “the cat”. In many languages of the world, however, speakers would also indicate the grammatical gender of the cat. In German, for instance, the gender of a noun is marked on the article so that a cat, regardless of its biological sex, is feminine, (die$_{\text{feminine}}$ Katze$_{\text{cat}}$), but a dog is masculine (der$_{\text{masculine}}$ Hund$_{\text{dog}}$). The German gender system actually comprises three genders, masculine (*der*), feminine (*die*), and neuter (*das*), and varying sources describe the distribution of each in the German lexicon as being either 50%, 30%, 20% respectively (Bauch, 1971, as quoted in Hohlfeld, 2006) or, based on the CELEX database and taking word frequency into account, 43%, 38%, and 19% (Schiller & Caramazza, 2003). What makes gender particularly difficult to learn in German is that its assignment, as can be seen from the example above, is relatively arbitrary. Unlike some gender systems such as Tamil, where nonhuman objects are classified as neuter while human objects are classified by their natural gender, there is no inherent connection between the biological gender and grammatical gender of a noun. I use the term “relatively arbitrary” to describe the German system because research now suggests that there is at least some systematicity to a noun’s gender assignment (cf., Bordag, Opitz, & Pechmann, 2006; Bordag & Pechmann, 2007; Schiller, Muente, Horemans, & Jansma, 2003; Schwichtenberg & Schiller, 2004), a point to which I will return shortly.

---

[2] As a Germanic language, English used to have a gender, as well as a case system, similar to German, but these disappeared by early Middle English (cf. Cambridge History XIX.4).

In many ways grammatical gender is a lexico-syntactic feature that is distinct from other language features. Research suggests that grammatical gender is processed differently from properties such as number, and violations in gender show heavier processing costs than number violations (Barber & Carreiras, 2005). Gender also provides a challenge to general assumptions of language learning. In L1 learning, there is limited evidence that even adult L1 speakers of a gender-inflecting language continue to have difficulty in assigning gender, in this case for phonologically ambiguously marked nouns (Bates, et al., 1995)[3]. The first half of this section reviews several monolingual models on how gender is represented in the mind, along with supporting empirical evidence. I then address the question of whether gender is represented the same way in the L2 and to what extent native-like representation is possible. Motivations for the current study within the framework of gender processing conclude this chapter.

2.3.1 Representation of Gender in the L1 Lexicon

Linguistic research addresses the question of the representation and organization of language in the mind, and several models capture these theories of organization. The overarching idea behind these models is that the mental lexicon stores words according to their phonological, syntactic, and semantic properties, as well as non-linguistic properties, and that these characteristics form relationships between words and even parts

---

[3] Although, rather than reflecting true difficulty with gender, it is possible that the class of nouns under investigation was an irregular form gradually regularizing (e.g., Pinker & Ullman, 2002). In L2 learning, gender seems particularly sensitive to the influences of the L1 and relatively insensitive to other factors, such as general L2 proficiency, showing persistent deficits even after the language learner has attained general fluency (e.g., Brun de Garavito & White, 2000; Sabourin, 2006).

of words. This organization is thought to influence the way we process language, as reflected in the speed at which we recognize, produce, or make decisions about language. Several theories exist about exactly how this organization and processing takes place and precisely when grammatical gender becomes available to the system.

The discussion here is limited to visual word recognition and comprehension, in view of the experiments that will follow (for a comparison of word recognition and word production, see Heim 2005). The starting point for each of these theories is the idea that in order to read a word, the reader must access the phonology, orthography, and meaning of a word. Comprehension then involves a word's syntactic and thematic functions (cf., Coltheart, 1978; Glushko, 1979; Seidenberg & McClelland, 1989). Under debate, and important to the current discussion, is how prior gender information presented in a gender marked article or adjective affects the processing of a subsequent word.

2.3.2 Models of Language Processing and Gender

Accessing a word in our lexicon is a complex process. From the moment that we see a word to the instance of retrieving it from our lexicon, our mind goes through many steps before a word is identified. And yet all these steps are rapidly completed: the skilled reader can process five words or more per second (Rayner & Pollatsek, 1987). Research in support of different models of reading often divides this lexical access process into two stages, a prelexical/lexical and a postlexical stage (see Hernandez, Bates, & Avila, 1995, for a review). The prelexical stage encompasses the processing which occurs before the word is recognized and which is largely automatic, without the need for reflection, while the postlexical stage involves processing after the lexical item has been accessed. This second stage is thought to be more controlled and under the influence of the reader (cf., Posner & Snyder, 1975, for more on the distinction between automatic and controlled processing). While two models within this framework, the Checking model and the Intralexical model, maintain a more modular (i.e., independent) view between these two stages, a third, the Interactive model, views the stages as interacting with one another. As a result, each model makes different proposals about when gender information from a previous article becomes available in the processing of a subsequent noun.

2.3.2.1 The Intralexical Model

The Intralexical Model for word processing draws analogies with word production models, such as those by Levelt and colleagues (e.g., Levelt, 1989; Levelt, Schriefers, Vorberg, Meyer, Pechmann, & Havinga, 1991; Level, Roelofs, & Meyer, 1999) and distinguishes between the retrieval of the so called "lemma", the abstract lexical entry and associated syntactic and semantic information, and the actual word form or lexeme which specifies the phonology (Fodor, 1983; Swinney, 1979; for reviews see Bates, Devescovi, Pizzamiglio, D'Amico & Hernandez, 1995; Frauenfelder & Tyler, 1987; Friederici & Jacobson, 1999). In this division, the grammatical gender of a noun is part of its lexical representation. The Intralexical Model makes similar adaptions for word recognition as Dell and colleagues have made for word production (e.g., Dell & Sheaghda, 1992). Significantly, unlike the Levelt model for word production, the intralexical model permits feed-back between the gender node and the lemma node so that gender information from the article becomes relevant to the processing of the noun pre-lexically or lexically. Reading a gender-marked article would activate the gender node, and this activation would spread to all nouns with the same gender. As a result, the gender of a preceding article could function to narrow the selection of following nouns, speeding the ultimate noun selection. In this way, gender information becomes available pre-lexically or lexically, while maintaining the modularity of the theory.

2.3.2.2 The Checking Model

The Checking Model by Friederici and Jacobson (1999) does not necessarily specify the architecture in the mental lexicon like the Intralexical model, but rather emphasizes a neurocognitive model based on the phases of processing (Friederici, 2002). Word processing moves from the identification of phonemes, to the identification of the word form, then the word category, the lemma and morphological information, semantic and syntactic integration, and finally processes of reanalysis and repair. As a modular theory, the Checking Model argues that the preceding gender information, such as that of an article, is not used to control lexical access. Instead, the gender information of the article stays active until after the noun is accessed and is then used in a syntactic congruency check against the lexical gender information of the noun in a post-lexical checking mechanism (see also Faussart, Jakubowitz, & Costes, 1999).

2.3.2.3 Interactive Models

Interactive models are based on connectionist views, rather than modular views of the lexicon. As a result of the nonmodularity, semantic or syntactic context information can interact with the lexical processing to reduce the possible candidates for selection (e.g., Bates, Devescovi, Hernandez, & Pizzamiglio, 1996; Bates, Elman & Li, 1994; Elman & McClelland, 1988; MacDonald, Pearlmutter & Seidenberg, 1994; MacWhinney, 1989; Rumelhart & McClelland, 1986). Many sources of information, from the sentence as a whole to the context of the discourse, can serve to activate and constrain possible words for selection, often before the reader encounters the actual word. In essence, the reader predicts the upcoming word based on the information at hand. In this model, gender, as presented in a preceding article or a gender-marked adjective to a noun, could serve to pre-activate a subgroup of nouns with the same gender, or inhibit a subgroup of nouns with a different gender. The pre-lexical nature of this model makes similar predictions to Intralexical models, and as a result the two are difficult to distinguish (cf., Friederici & Jacobson, 1999). However, both the Interactive and Intralexical model differ strongly from the post-lexical checking mechanism offered by the Checking model.

2.3.3 Empirical Support for the Representation of Gender

Compelling evidence for the unique status of grammatical gender in the mental lexicon comes in part from research by Barber and Carreiras (2005), who directly compared grammatical gender and grammatical number. On an intuitive level, one would suspect that gender agreement and number agreement would function similarly; however, there is good reason why gender is typically considered a feature of the lexical representation while number is considered a morphological property (see also Koester et al., 2004, and Tokowicz & MacWhinney, 2005, for further discussion of the dissociations in processing number and gender). In the ERP study of Barber and Carreiras, participants read Spanish word pairs (Experiment 1) and sentences (Experiment 2) with noun-adjective (e.g., $faro_{lighthouse}$-$alto_{high}$) or article-noun pairs (e.g., el_{the}-$piano_{piano}$), which were matched or mismatched for gender or number. The critical result for word pairs was that while there were no significant differences in the size or distribution of effects for gender versus number agreement in the LAN or the N400, gender agreement violations produced significantly longer latencies than number disagreement in the P3 component. The P3 has been implicated in indicating stimulus relevance to a task, and its production indicates completion of stimulus categorization (Donchin, 1979). Longer latencies in gender agreement thus indicate that participants took longer to register gender violations than number violations, suggesting different underlying processing strategies or mechanisms for gender and number agreement. In Experiment 2, word pairs embedded in sentences showed a significant difference between gender and number disagreement only in late phases of the P600, and not in the LAN or earlier phases of the P600, replicating their previous results (Barber & Carreiras, 2003). These findings suggest that gender disagreement results in costlier reanalysis processes than number disagreement. Barber and Carreiras argue that the lack of a conceptual relationship for gender in many languages and its resulting arbitrary assignment to a noun forms the basis for the purely lexical status of gender. Because of gender's status as a lexical feature, a checking mechanism would need to monitor both the syntactic integration and lexical access processes in order to repair a detected inconsistency. Only a single syntactic integration process would need to check number agreement since number has an autonomous representation during syntactic analysis, thus leading to less costly reanalysis for number agreement than for gender agreement.

Other research has further examined the time course over which grammatical gender information becomes available to lexical and syntactic processing. This research also supports the findings of a late reanalysis in gender assignment. Early research on prelexical effects suggested that a grammatical gender prime would activate all lexical entries that matched the prime in gender. This proposal was problematic as there is a large subset of lexical items matching any given gender, leading to a rather inefficient strategy if used to discriminate among lexical alternatives. Instead, Schriefers et al. (1998), following the arguments of Tanenhaus, Dell, and Carlson (1987), proposed that the processing of gender information occurs early, but later helps to reduce the number of lexical candidates activated by previous semantic information. In this account, gender primarily serves an inhibitory function, but could also reduce lexical-to-lexical priming effects when there is a gender disagreement, as between a target noun and a definite article. This account seems to integrate late gender processing models, which argue for initial coarse-grained processing, ignoring features like gender until later processing (Mitchell et al., 1995), and early parsing theories in which immediate processing of gender blocks gender-incongruent alternatives via lateral inhibition (Vosse & Kempen, 2000).

The Mitchell et al. (1995) model is based on findings with native Dutch speakers in which they fail to use grammatical gender to disambiguate a relative clause, both on-line and to some extent even off-line (Brysbaert & Mitchell, 1996; 2000). However, later research in Russian challenged this view (Akhutina et al., 1999; Sekerina & Pugach, 2005; van Berkum et al. 1999). As pointed out by Sekerina & Pugach (2005), the research by Brysbaert and Mitchell depended on a syntactic use of gender which has largely changed in present-day colloquial Dutch; gender is now rarely used in relative clauses on relative pronouns. Together with the results of other studies, the evidence would point to the probability that gender is used during syntactic parsing to disambiguate sentences.

Several studies have further investigated whether gender information interacts with semantic relatedness at later stages of sentence processing for native speakers of gender-marked languages (Friederici & Jacobsen, 1999; Guillelmon & Grosjean, 2001; Gunter et al., 2000). In an ERP study with monolingual German participants, Gunter et al. (2000) further confirmed the idea of early parallel activation and later interaction. Although there is debate concerning the interpretation of the components of the ERP record, typically the N400 effect is associated with lexical and semantic integration processing whereas syntactic violations are associated with the P600 (Brown & Hagoort, 1993; Osterhout & Holcomb, 1992). Gunter et al. showed that semantic and syntactic processes run in parallel at early stages, such that gender information and word meaning are available concurrently. This parallel processing interacted only at later stages: all nouns of low cloze probability elicited an N400 regardless of gender mismatch compared to nouns of high probability, while gender violations between the definite article and the noun evoked a LAN; these two effects did not interact. However, the two variables did interact during a later P600 component.

Recent child language research has also shown that children learning Spanish as a first language are able to use grammatical gender to establish reference in sentence interpretation as early as 34 to 42 months (Lew-Williams & Fernald, 2007). Using the same methodology, Lew-Williams and Fernald showed the same ability in native-Spanish speaking adults, suggesting that by the age of three, children already possess a highly developed gender processing system.

2.3.4 Representation of Gender in the L2

The question then arises whether L2 learners of a language with grammatical gender process gender in this same way as native speakers. In particular, two models of L2 learning introduced earlier in Chapter 1 make competing predictions about the ultimate ability of an L2 speaker to learn grammatical gender. According to the Full Transfer/Full Access Hypothesis (FTFA, Schwartz & Sprouse, 1994; 1996) the L2 learner transfers the entire L1 grammar as the starting-point of learning. Once the L1 grammar is transferred, learners have full access to Universal Grammar (UG) during the course of development to restructure parameters and to acquire even L2 structures not present in the L1. In the context of the FTFA, L2 learners could develop native-like representations of gender even as late learners and regardless of whether the L1 already had a gender representation or not (although gender in the L1 may facilitate L2 gender acquisition). On the other hand, the Failed Functional Features Hypothesis (FFH, Hawkins & Chan, 1997; Hawkins & Franceschina, 2004) maintains that certain features of a language can only be learned with great difficulty, if at all, after a certain critical period. According to this hypothesis, after the onset of puberty, L2 learners only have access to functional features already extant in their L1. As a result, L2 learners must approximate new features by using compensatory strategies such as explicit rule learning. L2 learners without representations for gender would never develop them, and would need to rely on explicit rules, if available, to assign gender to nouns.

Particularly a late syntactic process of “gender checking” for L1 speakers, as discussed by Friederici and Jacobsen (1999), may not be available in the same way to L2 learners. However, the behavioral study of Taraban and Kempe (1999) questions the viability of such checking mechanisms, and computational models in general, suggesting instead that a cue-based/connectionist model may afford a more parsimonious mechanism for both L1 and L2 gender processing. While not closed to the possibility that both gender-marked lexical representations, as well as heuristics for processing gender, could be used, they question how individuals would choose which method to use (cf. discussion on p. 121). Taraban and Kempe argue that particularly for L2 learners, a checking-mechanism makes little sense if they do not know the gender to be checked in the first place. Their data showed that both L1 and L2 speakers of Russian relied on grammatically inflected adjectives to help disambiguate the gender of nouns and to choose appropriate verb forms in a forced-choice task. However, only L2 speakers showed faster naming latencies for reading opaque (i.e., ambiguous) nouns in comparison to transparent nouns when there was a disambiguating adjective. The results suggest that L1 speakers’ performance was already at ceiling, reflecting learning levels that were at asymptote. L2 performance, however, revealed sensitivity to noun-endings. Based on connectionist modeling, Taraban and Kempe suggest that L1 and L2 speakers of Russian were on a different point of the same continuum in learning. If L1 and L2 gender processing are on the same continuum, a checking mechanism would not support the data as readily because such a mechanism should not be sensitive to cues such as the reliability of noun endings as reflected in the L2 data.

Other factors that underlie the assignment of grammatical gender in L2 learners are noun frequency (Sabourin, 2006), and congruency effects (cf. Guillelmon & Grosjean, 2001). Guillelmon & Grosjean found that there were strong grammatical congruency and incongruency effects for monolinguals and early bilinguals in an auditory naming task in which participants had to name the final noun of an auditorily presented phrase (e.g., la jolie glace – the pretty ice). Late bilinguals were insensitive to congruency, in that RTs were the same regardless of whether the determiner and adjective matched or mismatched. Based on these results, Guillelmon and Grosjean were able to rule out speed of response, production skills, and language proficiency as possible explanatory factors. This research suggests that at least late bilinguals engage a different process to compute gender.

However, behavioral data may not always reflect on-line processing and there is the possibility that although learners are not able to produce gender the way a native speaker would, their neural processing may still reflect sensitivity to gender. Tokowicz and MacWhinney (2005) specifically exploited ERPs to test L2 learners because the more temporally fine-grained resolution of the ERP record allows the identification of distinct components in processing over time. They postulated that ERP methodology would be sensitive to implicit knowledge that is often not reflected in behavioral data. Their ERP results showed that, indeed, native English speaking L2 learners of Spanish implicitly processed gender disagreement, even though these same participants were at chance making these discriminations on an offline grammaticality judgment task. When these L2 learners then underwent training in a pilot study, they showed marked improvement in their explicit knowledge of gender agreement. This suggests that L2 learners can indeed "learn" grammatical gender and may simply be at an earlier stage on a continuum of learning. This is precisely what Taraban and Kempe (1999) suggested in their model. They were able to track the emergence of gender competence as a function of experience using modeling, revealing that L2 and L1 users are at different points on the same learning curve in the emergence of grammar[4].

Sabourin (2006) and Taraban and Kempe (1999) also suggest that the relationship between the bilingual's two languages plays a role in the initial sensitivity to grammatical gender. In a study investigating L1 transfer effects in L2 learning, Sabourin (2006) showed that the nature of the L2 learner's L1 was the most influential factor in determining the success with which the gender system was learned and used. This achievement was above and beyond general language proficiency, showing that the knowledge and use of gender stands separate from general syntactic proficiency. While L2 learners of Dutch from various language backgrounds could successfully complete a

---

[4] In contrast, Hahne and Friederici (2001), discussed earlier in this chapter, showed an absence of the P600 for late bilinguals and a delayed N400 effect signifying that bilinguals process the L2 both slower and differently than monolingual counterparts. Since the bilingual speakers in the study by Hahne and Friederici were Japanese-German speakers, the differences in the two studies could be due to the degree of similarity between the L1 and the L2. In fact, Tokowicz and MacWhinney found that for determiner number violations, English-Spanish bilinguals were not sensitive to these grammatical violations, postulating that a mismatch between the L1 and L2 structures could have caused interference.

gender assignment task, performance on a gender agreement task correlated strongly with the language background: language learners with a similar gender system to Dutch (German) had the highest scores, followed by learners with a different gender system (French). Learners who were native speakers of English, a language that does not overtly mark grammatical gender, performed at chance for gender agreement even though they performed relatively well on the gender assignment task. While they seemed to "know" the correct gender as reflected in their high identification scores, they were not able to use gender in the agreement task. The results suggest that language pairings may play a crucial role in determining the ability to transfer a grammatical system from L1 to L2. Other research also suggests that processing resources may affect language transfer so that native speakers of a language without a particular construction in their L1 needed in the L2 may require additional cognitive processing resources such as working memory to acquire the ability to use that construction on-line (Hoshino et al., in revision).

There is also evidence that both L1 and L2 users of languages with genders use frequency heuristics to help select appropriate gender markers (see Sabourin, 2006, for a discussion of this). Sabourin found that both L1 and L2 speakers use a default strategy of assigning the more frequently used gender in situations where gender assignment is more ambiguous. Since both L1 and L2 speakers use this heuristic, this would suggest that underlying language processing between native and L2 learners may not be that different after all, or that they at least have access to the same implicit inferencing mechanism.

Other research further points to the importance of morphophonological, semantic, and syntactic cues (that is, noun endings, natural gender, and agreement cues respectively) in assigning gender (Oliphant, 1998; Bordag & Pechmann, 2008). Research such as Oliphant (1998) and Tokowicz and MacWhinney (2005) focuses on languages where gender assignment follows extremely regular phonological patterns (e.g., words in Italian as well as Spanish ending in /a/ are typically feminine and /o/ are masculine)[5]. Gender congruence between article and noun are thus fairly transparent, making this type of system easier for L2 learners to acquire than that of a language, such as German, where phonological patterns are not as regular or widespread. Where phonological patterns do exist in German, learners seem to be quick to pick up on these cues in order to boot-strap their way into the gender system (e.g., Bordag, Opitz, & Pechmann, 2006).

Recent research in the learning of grammatical gender in children has noted general cross-linguistic differences, with Romance systems, such as Spanish, typically being easier to acquire than Germanic languages (cf. Karmiloff-Smith, 1978; Kuchenbrandt, 2008; Kupisch, 2007; Lleó & Demuth, 1999; Möhring, 2001), although German children do master gender by around the age of 3 (Mills, 1986). Both the transparency of the assignment rules as well as the saliency of the gender markings seem to come into play in determining the ease with which the system can be learned (see also Bates & MacWhinney, 1989, and Slobin & Bever, 1982, for the relevance of cue salience in language learning). For instance, a majority of the Spanish lexicon is made up of regular nouns in which nouns ending in *–o* are masculine and nouns ending in *–a* are feminine (cf. Harris, 1996). In contrast, formal properties of the German noun correlate rather weakly with a noun's gender class. However, those assignment rules that are

---

[5] One could also argue that the ERP results in Tokowicz and MacWhinney showing sensitivity to gender violations in early learners might in fact be due to phonological patterns which are relatively easy to detect and could therefore have nothing to do with the actual understanding of the gender system.

transparent do seem particularly salient, even to native speakers. According to Schwichtenberg and Schiller (2003), native speakers of German are indeed sensitive to semantic gender assignment regularities. Participants presented with categories and gender-marked pseudo-words preferentially selected pseudo-words marked with a gender associated with the given category (e.g., predators are typically masculine). Hofmann (2005), in an ERP experiment, also showed that native German speakers are sensitive to phonology and derivational morphology in gender assignment in addition to semantics. Participants were presented with a German noun and one of three gender adjectives (*weiblich* – feminine, *männlich* – masculine, or *sächlich* – neuter) and had to decide whether the gender pairing was correct or not. Results revealed an increased sensitivity to derivational-morphology over semantic gender, and increased sensitivity to semantic compared to phonological gender in guiding gender decisions.

Kuchenbrandt (2008) compared the learning of grammatical gender in monolingual Spanish and monolingual German speaking children with balanced Spanish-German bilingual children. Data collection comprised audio recordings made in a naturalistic setting between the ages of 1;02 (years; months) and 2;03. Both the bilingual Spanish-German children and the monolingual German children showed a delay in learning the phonological preconditions of gender marking in German in comparison to the learning of gender in Spanish by their monolingual Spanish counterparts. The bilingual group consistently showed a delay in Spanish compared to the Spanish monolingual counter-parts, but an acceleration in German in comparison to native German children (See also Kupisch, 2007, for similar results). It thus stands to reason that if children learning German as a native language show certain delays in gender learning by comparison with peers learning a Romance gender system, then it is plausible that late L2 learners of German could also show increased difficulty in learning grammatical gender compared to late L2 learners of languages such as Spanish.

A difficulty in deciding whether gender processing in L2 learners will ever approximate that of native speakers is that the ERP studies reviewed here, while immensely helpful in disentangling the time-course of processing, have not controlled for the level of language proficiency. In fact, the bilingual studies have by and large used L2 speakers of intermediate proficiency. It is hard to know then whether L2 users who behaviorally display near-native ability in assigning gender would also show a time-course of processing similar to native speakers (e.g., Perani et al. 1998) or, alternately, whether a behavioral insensitivity to gender might show electrophysiological evidence for sensitivity (Tokowicz & MacWhinney, 2005). L1 ERP studies typically show the influence of gender in later syntactic processing. Interestingly, effects of bilingualism emerge at this later processing point (e.g., the P600), implying that bilinguals may have particular difficulties with syntactic integration or reanalysis. It stands to reason that if connectionist models of gender processing are correct, bilinguals should show an increasing ability to engage in these late syntactic processes the more they are exposed to a language. The present study thus further investigates the representation of grammatical gender and its implications for their accessibility in the L2. While previous work has looked at high (Sabourin, 2006) and low (Tokowicz & MacWhinney, 2005) proficiency learners, the current study investigates a wider range of proficiency, including both a set of high and low proficiency learners. By looking at German, in which gender assignment is less transparent than in languages such as Spanish (cf. Tokowicz & MacWhinney, 2005), the role of proficiency in modulating the acquisition of an arguably more difficult-to-learn gender system is examined. Furthermore, the current study goes beyond looking at gender knowledge and additionally addresses lexical access of gender in L2 learners, providing further evidence for whether L2 learners of a gendered language can use gender on-line.

Chapter 3

General Directions and Methodologies

The review of the literature in Chapter 2 provides a picture of the complexity of L2 learning. While some of the behavioral and neuroimaging research suggests that proficiency is key in modulating constraints on L2 acquisition (e.g., Abutalebi et al., 2005; Perani et al., 1998), other research provides equally compelling evidence for the role of age of acquisition and L1-L2 pairings (Birdsong & Molis, 2001). A possible reason for the apparent discrepancy in the conclusions of these studies is that different studies have examined different aspects of the semantic/syntactic interface. The focus on grammatical gender and morphology in the present work was chosen deliberately because these are the aspects of language processing that have often been reported to be the most difficult for L2 learners to acquire. The current dissertation study addresses these issues using native speakers of English learning German and native speakers of German learning English. The language pairing of German and English is ideal because while English is a Germanic language, it no longer uses a grammatical gender system. By examining a structure that does not exist in the English-German bilingual's L1, the study can provide further needed insights into issues of language learning as well as help explain how bilinguals achieve language control and modulate potential cross-language activation.

3.1 Goals of the Present Research

The current study further examines the role of gender in bilingual language processing using both cognitive and neurocognitive measures. The aim is to identify the factors that constrain access to aspects of the grammar and semantics in the L2 and then to localize their effects over the time course of processing. Research in the monolingual domain has not been able to fully determine when grammatical gender becomes available for processing during lexical access. The results of the current study can potentially adjudicate between alternative models of gender processing, while clarifying the nature of L2 gender representation.

To further investigate these ideas, the study examined German grammatical gender in simple and compound nouns. German has a gender system in which nouns are marked for masculine, feminine, or neuter gender. In simple noun processing, as reviewed in the previous chapter, gender assignment is relatively arbitrary (e.g., $\text{die}_{\text{feminine}}$ $\text{Katze}_{\text{cat}}$, which indicates either a male or female cat), although there are some phonological rules which guide selection (in the example above, nouns ending in –e typically take the feminine gender). How do learners of German whose native language does not have a comparable gender system learn an arbitrary assignment?

Gender processing in compound nouns is more complex, but also more systematic. While gender assignment in compounds still follows the same arbitrary system, if a person knows the gender of each of the constituents, then assignment of gender to the entire compound is relatively easy: Compound nouns take their gender from the final noun so that while “Wein” is masculine and “Glas” is neuter, the compound noun “Weinglas” is neuter. Previous ERP research has shown that native speakers of German are sensitive to compound nouns that show an incongruence in the genders of their constituents, such as in the example of “Weinglas,” leading to the conclusion that at least for native speakers, compound nouns are processed as decomposed units (Koester et al., 2004). This raises the question of whether L2 learners of German show this same sensitivity to gender incongruence and whether L2 proficiency modulates the process of decomposition. If L2 learners are not sensitive to grammatical gender as they process the L2 on-line, then they should be indifferent to the congruence of the gender and the noun. If they are sensitive to gender, then gender/noun congruence should matter, with different response latencies for compounds in which the constituents’ gender match compared to compounds in which there is an incongruence. A related question, however, is whether the L2 learners decompose at all. One possibility is that they decompose compounds like native speakers, but are not sensitive to the gender mismatch. Another possibility is that they treat compounds like single lexical units, in which case they would also not be sensitive to the mismatch.

The current chapter lays out the framework for the experiments conducted on gender processing in L2 learners. In an initial step, learners with varying degrees of proficiency were tested on their knowledge of grammatical gender for simple nouns. The experiment was then adapted for use with ERPs in order to gain a more fine-grained temporal analysis to investigate the possibility that the implicit processing of grammatical gender may not be reflected in behavioral or off-line tasks (e.g., Tokowicz and MacWhinney, 2005). The third experiment examined the issue of grammatical gender in L2 morphological processing to determine whether L2 learners of German represent the two noun components of a compound noun as separate lexical items or as a single lexical item. The experiments involved not only different types of language pairings (native English learners of German and native German learners of English) but also intermediate and advanced proficiency speakers for each language pair. This allowed us to examine the effect of the development of proficiency on issues of language transfer. Since all three experiments used the same translation recognition task, I first present a general overview of the methods for each of the experiments here. A more detailed discussion of compounding, relevant specifically to the predictions of Experiment 3, will be reserved for Chapter 5 where that experiment is described.

3.2 General Method

Each of the experiments used a translation recognition paradigm to address grammatical gender and compound issues in order to further evaluate lexical access in the mental lexicon, as well as the architecture used in decomposing morphologically complex words. In translation recognition, a word is presented in one language and followed by a word in the other language. The bilingual must decide whether the second word is the correct translation of the first word (De Groot, 1992). In the variant used in the present experiments, the critical comparison is between word pairs that are not translation equivalents and their controls matched on word length and frequency. The speed with which the participant is able to reject a nontranslation word pair depends on the amount of interference caused by form and meaning overlap between the two words. Learners at different levels of proficiency may be differentially sensitive to form and meaning overlap, as well as to gender incongruence. Previous research has shown that translation recognition is sensitive to developmental changes in reliance on the L1 form (Ferres et al., 2006; Sunderman & Kroll, 2006; Talamas, Kroll, & Dufour, 1999). Unlike production tasks which require greater L2 proficiency for successful completion, translation recognition is a receptive task, which even less proficient L2 learners are able to perform at a relatively high level of accuracy. As a result, we were able to test participants of lower proficiency in addition to higher proficiency bilinguals, allowing for a wider range of proficiency among the present sample of participants. These characteristics made translation recognition ideal to further investigate the role of proficiency in modulating grammatical access. Translation recognition additionally avoids the complication in many production tasks where pictures are named of using only concrete nouns; in this way more abstract nouns could be tested with lower-proficiency participants, tapping into a broader spectrum of the mental lexicon. An additional advantage specifically for the ERP study was that using a receptive task reduced artifacts caused by muscle movements, a particular challenge in ERP production studies. Each of the on-line experiments was followed by an off-line gender assignment task of the critical items (e.g., Jiang, 2004; Tokowicz & MacWhinney, 2005) in order to assess off-line explicit processing of gender and the role of task-demands (McLaughlin et al., 2004).

Translation recognition paradigms use either forward or backward translation orders. In backward translation, participants first see a word in their L2, followed by a word in their L1. In forward translation, participants first see the word in the L1, followed by the second word in the L2. Previous research has either used a combination of forward and backward translation (e.g., de Groot & Comijs, 1995; de Groot, Delmaar, & Lupker, 2000; Talamas et al., 1999) or backward translation exclusively (Sunderman & Kroll, 2006). Experiment 1 used only forward translation from English to German in order to allow preparation of the gendered article in German for L1 English speakers. The order of presentation (i.e., English first, German second) was not changed for the native German speakers, however, in order to keep any gender preparation in translation consistent across the language groups. As a result, native English speakers engaged in forward translation, while native German speakers completed backward translation. In addition, each noun was preceded with an article to explicitly activate gender in word recognition. Participants therefore saw sequences such as THE CAT – DIE KATZE and then had to respond whether the second noun phrase was a correct translation of the first noun phrase. All three experiments used a version of the task in which participants respond "yes" to correct translations and "no" for incorrect translations via a button press on a keyboard or button box. In all cases, the critical trials were the "no" trials. In Experiment 2 the paradigm was adapted for EEG, thus enabling ERP data collection. In ERP studies, one concern is that movements, such as eye movements from reading and eyeblinks, cause artifacts in recording which obscure the data. To minimize movement artifacts, the timing of the sequence was modified for the ERP study to be detailed in Chapter 3. The compound study, Experiment 3, used the same procedure and timing as Experiment 1.

A subset of the materials included cognates, a situation common for a language pairing such as English and German with common linguistic roots. In all three experiments, rather than attempting to limit items to non-cognate stimuli, cognates were included as a subset of the critical items. Cognates were matched to noncognate controls on lexical factors, such as frequency, familiarity, and word length, in order to assess the effects of cognates status and cross-language lexical activation. Because of the close link of cognates across languages, and the natural occurrence of cognates within each language, they can provide a window into L1 transfer issues. Particularly in a language pairing (English-German) in which the L1 does not have grammatical gender, cognate processing could potentially reveal an increased difficulty in acquiring sensitivity to gender for this subset of words. Similar to other recent bilingual studies of grammatical gender, greater difficulties in gender processing were anticipated for cognates than noncognates (Lemhöfer, 2006).

3.2.1 Participants

Participants were recruited from two different groups: (1) native English learners of German, and (2) native German learners of English. Since German-English learners drawn from a German university setting are typically already more proficient in their L2 than their US counterparts, experiments in Germany focused on relatively high proficiency late L2 learners. To assess L2 proficiency, participants completed a language experience questionnaire that included self-ratings of L2 proficiency. In addition to the language background questionnaire, participants also completed online measures of cognitive performance, including memory span (using the operation span task) in the participant's L1 (Turner & Engle, 1989; Tokowicz et al., 2004), the Simon task, a measure of executive control (e.g., Bialystok et al., 2004; Lu & Proctor, 1995; Simon & Rudell, 1967), and a simple picture naming task in both L1 and L2 to assess language proficiency and dominance (e.g., Jared & Kroll, 2001). Each of these tasks is described in more detail below. Using these ancillary tasks, participant groups could be matched on cognitive resources, and proficiency differences between groups across experiments could be controlled. Individual difference measures in cognitive abilities were important not only to match groups but also to evaluate processing resources needed for L2 learners to acquire and implement a language construction which does not exist in their L1. Their ability to acquire gender in German and use it online may be dependent on available processing resources (e.g., see Hoshino et al., under revision, for a similar argument about the acquisition of subject-verb agreement).

3.2.2 Experiment 1: Simple Nouns

In Experiment 1, participants judged translations of simple nouns such as THE CAT – DIE KATZE. As seen in the example below, the presence or absence of the correct translation was manipulated and whether the gender of the German article agreed or conflicted with the simple noun. Correct and incorrect translations were presented with gender matches and mismatches to the target translation gender. For the correct translations to which participants responded *yes*, fillers were developed which were frequency and length matched for the critical words. There were three critical "no" conditions:

1) German items which did not match the critical English item in either gender or the translation e.g., THE SCARF (der Schal) – DAS BRETT (the board)
2) German items which matched the critical English item in gender but not the translation e.g., THE SCARF (der Schal) – DER KNOPF (the button)

3) German items which matched the critical English item in the translation, but not in the gender e.g., THE SCARF (der Schal) – *DAS SCHAL (the scarf)

Note that in this third condition, the presented German article and noun are ungrammatical, whereas in Conditions 1 and 2, the German article and noun pair are grammatical, although they are the wrong translation of the English word.

As reviewed in Chapter 2, there seems to be a discrepancy between off-line and on-line performance. While L2 learners seem to master grammatical gender in off-line assignment tasks, they show persistent difficulties with more on-line processing. The logic of the Experiment 1 was to determine whether a receptive task, such as translation recognition, would reveal greater sensitivity to grammatical gender than previously reported for productive tasks. Another goal of Experiment 1 was to determine whether native-like L2 gender processing is possible for late L2 learners. While several lines of research point to the role of language proficiency in acquiring grammatical gender systems, research such as that done by Guillelmon and Grosjean (2001) suggests that late bilinguals as compared to early bilinguals process grammatical gender in a completely different manner, which is not modulated by their level of proficiency. Similarly, Silverberg and Samuel (2004) and Kotz and Elston-Güttler (2004) showed that language pairing and age of acquisition constrain access to word meaning in the L2. It is therefore likely that even highly proficient late bilinguals will not process gender mismatches in the same way as native German speakers.

In contrast to these findings, Sabourin (2006) showed that high proficiency English-Dutch bilinguals were able to successfully complete a gender assignment task, but only performed at chance on a gender agreement task. Depending on whether the translation recognition task is most similar to a gender assignment or gender agreement task, it is possible that native English speakers who had achieved a high level of proficiency in German as the L2 would show sensitivity to gender mismatches. If proficiency modulates language transfer, then high proficiency English-German learners might be able to process gender similar to the German-English learners.

The pattern of results may also depend on available cognitive resources in addition to language proficiency. While research such as Clahsen and Felser's (2006) questions the role of certain individual differences such as working memory (WM), other research has implicated a role for WM in language processing. Michael and Gollan (2005) specifically suggest that WM may be important in a bilingual's ability to inhibit or suppress the unintended language of processing. The research of Hoshino, Dussias, and Kroll (under revision) also suggests a curious dissociation of WM in language processing. In a study examining subject-verb agreement, they found that higher WM span monolinguals were more sensitive to distributive number (e.g., "the drawing on the posters") than lower span monolinguals. No such correlations with WM occurred for bilinguals. However, bilinguals as a whole had significantly higher reading spans than their monolingual counterparts. So while span did not seem to make a difference in how bilinguals processed distributive number, the very fact of bilingualism seemed to have impacted their WM resources (see also Kroll, Michael, Tokowicz, & Dufour, 2002). I know of no previous research that has investigated the role of cognitive resources in modulating language transfer effects in this context. English-German bilinguals with higher cognitive resources may as a result be more likely to acquire structures in the L2 that do not exist in the L1.

3.2.3 Experiment 2: Simple Nouns in ERPs

A particular challenge in designing bilingual studies is the fact that behavioral tasks often have difficulty delineating the exact time course of language processing. Recent evidence from ERP studies even points to a dissociation between behavioral and ERP results, suggesting a distinction between implicit and explicit grammatical processing (Kotz, Holcomb, & Osterhout, 2007; McLaughlin, Inoue, & Loveless, 2000; Tokowicz & MacWhinney, 2005; Weber-Fox & Neville, 1996). To address this problem, Guo, Misra, Tam, and Kroll (2008) have adapted the translation recognition task for use with ERPs to be able to compare behavioral and ERP measures in this paradigm. To my knowledge, translation recognition has only been used infrequently with ERPs (Brenders, van Hell, & Dijkstra, 2007; Vigil-Colet, Pérez-Ollé, & García-Albea, 2000), thus providing converging evidence for the behavioral studies, as well as giving a more detailed picture of the time course over which gender is processed. In Experiment 2, participants were presented with word pairs like those presented in Experiment 1, focusing on grammatical gender assignment to single constituent nouns.

Previous ERP research on gender agreement with monolinguals has shown that incongruence in syntactic gender agreement elicits a left-anterior negativity (LAN) (Barber & Carreiras, 2005; Gunter, Friederici, & Schriefers, 2000; Koester et al., 2004). I therefore expected that if L2 learners register gender incongruence, they should show a LAN (see also Zwitserlood, 1994). Gender violations within sentence contexts have also elicited a P600 (Gunter et al., 2000; Hagoort & Brown, 1999), but it is unlikely that this effect can be captured in simple article-word presentations as used in the present studies.

Translation mismatches, on the other hand, should tap into semantic processes reflected by an N400 (e.g., Barber & Carreiras, 2003), which is likely to be modulated by language proficiency, such that more proficient participants exhibit larger N400 effects than less proficient participants (e.g., Kotz & Elston-Güttler, 2004). The N400 and LAN are likely to interact, so that there will be a pair of effects, one showing decreased N400 amplitude for correct translations and another a LAN for gender mismatches. For a pairing such as "THE CAT – DIE$_{feminine}$ PUPPE$_{doll}$", there may be an absence of a LAN since gender matches across translation, but there should be an increased N400 because of the translation mismatch. In contrast, in a pairing such as "THE CAT – DAS$_{neuter}$ KATZE$_{cat}$", there should be a LAN because of the mismatch in gender, but a reduced N400 for the correct translation. Using translation recognition, we can therefore take advantage of the fact that the LAN and N400 tap into different syntactic and semantic processes to further determine how these two processes interact.

If gender sensitivity is modulated by language proficiency, then like the predictions for the behavioral experiments, we would expect increasing sensitivity to these effects such that lower proficiency English-German bilinguals show a decreased effect on the N400 and LAN as compared to higher proficiency English-German bilinguals. There may also be a dissociation in behavioral and ERP effects similar to Tokowicz and MacWhinney (2005), whereby learners show a seeming insensitivity to gender effects in the behavioral studies but reveal sensitivity to gender incongruence in the ERP study.

3.2.4 Experiments 3: Compound Nouns

If less proficient L2 learners of German are not as sensitive to grammatical gender, then it is likely that they will judge all correct translations as “correct” even in the presence of gender incongruence because they will focus only on the translation of the base noun. It is also possible that they will judge the translation correct as long as the gender agrees with either the first or the second constituent, but will not have internalized the rule that the final constituent determines the gender of the compound. With increasing proficiency, learners may become increasingly sensitive to this rule. Thus, the performance of proficient but late English-German learners would be expected to resemble that of native German speakers.

Previous research with monolingual English speakers processing compounds suggests that the final constituent of a compound acts as an access code to the full meaning of the compound, where initial or concurrent attention is focused on the final constituent (Juhasz et al., 2003). As a result, this natural attention to the final part of the constituent in English might facilitate correct retrieval of the compound gender in German, even if the L2 learners haven’t internalized the rule that the final constituent in German defines the compound gender. If L2 learners are aware of the gender of this final constituent, they will then correctly assign the gender to the entire compound. Accuracy results from a gender assignment task will serve as baseline comparison to determine whether participants knew the gender of the individual constituents.

A possible confound was the possibility that the relative frequency of the compound itself could influence whether morphological decomposition takes place with compounds in bilinguals. For instance, compounds of high frequency may encourage whole-word processing of the compound and gender since learners would have encountered the compound and its associated gender more frequently. In contrast, low frequency compounds may encourage decomposition of the compound in order to extract its meaning based on the constituents. This decomposition would be reflected in an increased sensitivity to grammatical incongruities across constituents, provided that the participants are sensitive to grammatical gender in the first place. Therefore, items were chosen from intermediate textbooks of German over a range of frequencies, and individual and overall constituent frequency was therefore controlled across conditions, in order to minimize the influence of item frequency.

3.3 Proficiency Measures

3.3.1 Gender Assignment Task

3.3.1.1 Materials and Procedure

After completing the translation recognition task, participants completed a computer-based gender assignment task in which they were presented with the German simple nouns and/or constituents of the compound nouns that made up the critical items in Experiments 1 and 2. German nouns were presented as black letters on a white background in Courrier New, 26 size font, and participants selected one of three keyboard keys to indicate the gender of the noun ("c" for "der", "b" for "die", or "m" for das). After presentation of a fixation sign (+) for 250 ms, participants were presented with the German word. Words stayed on the screen until participants made a decision or timed-out after 5000ms. Participants were told to make their selections as quickly and accurately as possible.

3.3.1.2 Data Analysis

Responses in which participants chose the wrong gender were excluded from reaction time and accuracy analyses as errors. Items with naming latencies below 300 ms and above 3000 ms and those deviating two and a half *SDs* from a participant's mean were also excluded.

The task served two purposes. First, it provided an approximation of their actual gender knowledge of the critical items, apart from the translation task. Because of the longer time-out, and even though reaction times were measured in addition to task accuracy, the task was more similar to off-line gender knowledge tasks used in other experiments (e.g, Jiang, 2004; Tokowicz & MacWhinney, 2005). In this way, more explicit processing of gender and the role of task-demands could be assessed (McLaughlin et al., 2004). Furthermore, the assignment task allowed us to perform post-hoc analyses on the translation recognition data, examining the pattern of response for data on which participants explicitly knew the gender, as reflected by their performance on the gender assignment task.

3.3.2 Picture Naming Task

As a measure of proficiency independent of the critical translation recognition task, participants were asked to name pictures in both their L1 and their L2. Pictures were chosen on a range of difficulties based on naming accuracy data from a previous L2 German picture naming experiment (pilot data for Misra, Guo, Bobb & Kroll, in preparation)

3.3.2.1 Materials and Procedure

In order to gain a closer to independent measure of proficiency in each of the languages, participants first named pictures in their L1 and then in their L2, as activation of the L2 before L1 is known to cause naming delays in the L1 (e.g., Jared & Kroll, 2001). Participants named 30 pictures in each language, using only the bare noun, and these sets were matched on English character and syllable length, English naming accuracy, German character and syllable length, German naming accuracy, English KF frequency, English CELEX frequency, German CELEX frequency, and English imageability across blocks/languages ($p > .05$). The language of naming for each set was counter-balanced across participants (but not the order of language presentation, as indicated above). An additional 15 pictures served as practice trials before each block. Naming was recorded using a digital recorder and coded for accuracy after the experiment. Please see Appendix A for the complete set of items.

Participants received written instructions on the computer screen which were reiterated by the experimenter. Their task was to name the pictures as bare nouns as quickly and accurately as possible. On each trial, participants were presented with a fixation point (+) at the center of the screen. They were asked to press the space bar of a keyboard to make the fixation sign disappear, allowing trials to proceed self-paced. The fixation was followed by a 500 ms blank interval, after which time the target picture appeared. The picture remained on the screen until the participant responded or timed out after 5000 ms, after which the fixation point reappeared and the next trial started. Participants named the first set of practice and 30 critical pictures in English, after which they were given a short break before continuing the next practice set and 30 critical pictures in German. If participants did not know the name of the picture, they were instructed to respond “no” “I don’t know” “nein” (no) or “weiß nicht” (don’t know).

3.3.2.2 Data Analyses

Responses in which pictures were incorrectly named or where responses were repaired were excluded from reaction time analyses as errors. Items with naming latencies below 300 ms and above 3000 ms and those deviating two and a half *SDs* from a participant’s mean were also excluded. Voicekey errors leading to repetitions of naming were excluded from reaction time analyses but were included for accuracy analyses. Technical errors causing a failure to name were treated as missing data.

Response accuracy was initially judged according to conservative and liberal criteria. Conservative criteria required exact production of the expected picture name whereas liberal criteria allowed some variation in naming to accommodate synonyms (e.g., “plane” for “airplane”). However, since picture naming served as a measure of proficiency, the conservative analyses will be presented here.

3.3.3 Language History Questionnaire

We assessed participants’ language experience and exposure using a questionnaire. Participants provided information about their native language, languages spoken at home, age of first exposure to the languages they knew, the length of any stays abroad, as well as a self-assessment of their L1 and L2 language capabilities in reading, writing, speaking, and listening comprehension. Self-ratings were made on a scale of 1 to 10 (1 being not proficient and 10 being very proficient). The full questionnaire is provided in Appendix B.

3.4 Individual Difference Measures

3.4.1 The Simon Task

The Simon task has been used extensively in the work of Bialystok and colleagues to assess the cognitive consequences of bilingualism (e.g., Bialystok, 2001; Bialystok, Craik, Grady, Chau, Ishii, Gunji, et al., 2005; Bialystok, Craik, Klein, Viswanathan, 2004). In the variation of the task used in the current study, participants saw either a red or blue square on a computer screen which appeared either to the left of fixation, at fixation, or to the right of fixation. Regardless of the square’s location, if the square was red, the participant pushed the tab key on the left-hand side of the keyboard, and if it was blue, the backslash key on the right-hand side of the keyboard. The task required inhibition of any prepotent response related to location and focus on the color of the square. In congruent trials, the square appeared on the same side as the correct button response (e.g., a red square appearing left that required a response with the left hand), while in incongruent trials, the square appeared on the opposite side of the button response (e.g., a blue square appearing left that required a response with the right hand). In central trials, the square appeared at fixation. The so-called Simon Effect, or measure of inhibitory control, was measured by subtracting RTs on correct congruent trials from correct incongruent trials. RTs on incongruent trials tend to be longer than congruent trials due to inhibition of the prepotent response to location (cf. Simon & Rudell, 1967).

3.4.1.1 Materials and Procedure

For each trial, a fixation sign appeared at the center of the screen for 350 ms, followed by a blank screen for 150 ms. A square (28 x 28 pixels), either red or blue, was presented either at center or 2° right or left of the center of the screen. The square remained on the screen until the participant responded or for 2000 ms. If the response was correct, a new trial started, after an interval of 850 ms. If the trial was incorrect, the word “ERROR” was presented in the center of the screen for 1500 ms, followed by an interval of 850 ms, before the next trial began.

There were three experimental blocks with seven trials for each condition—2 (colors) x 3 (location) for a total of 42 trials in each block and 126 trials in the task. The order of trials was randomized.

3.4.1.2 Data Analysis

Trials with incorrect responses or responses over 1500 ms were considered errors and excluded from both reaction time and accuracy analyses. Trials immediately following an error were also excluded from analyses. For each of the three conditions (congruent, incongruent, and central) mean RT and accuracy were calculated. The Simon effect was calculated by subtracting the mean congruent RT from the mean incongruent RT.

3.4.2 The Operation Span Task

The Operation-Span task, or O-span, provided a non-language specific measure of cognitive resources (La Pointe & Engle, 1990; Turner & Engle, 1989) and has several statistical advantages over other measures of working memory, including a high internal and test-retest reliability (e.g., Klein & Fiss, 1999).

3.4.2.1 Materials and Procedure

In the O-span task, participants read simple mathematical equations (e.g., (2*2) – 1 = 3) presented on a computer screen and had to decide whether the equation was correct or incorrect. As soon as a decision was made via button press, a word appeared on the screen. Participants were told to remember the word. After a set of two to six operation and word strings, the word *RECALL* appeared on the screen, and participants had to then recall the list of remembered words. 60 equations and 60 words were taken from Tokowicz, Michael, and Kroll (2004), which were based on Turner and Engle (1989), and were translated into German for use with the German-English bilinguals. As a result, participant groups completed O-span in their L1. Items in English and German were matched across each set size on English and German character length, German syllable length, German frequency (CELEX), German Leipzig frequency, English log frequency, English familiarity, and English imageability. Differences between sets were not significant (p > .05). Please see Appendix C for the full set of materials.

On each trial, participants initiated the trial by pressing either the “yes” button (TAB) or the “no” button (BACKSLASH). A fixation point (+) appeared at the center of the screen for 1000 ms and was then replaced by an equation. Participants were then to judge the accuracy of the equation by pressing either the “yes” or the “no” button. After the participant responded or after 3750 ms, a word appeared in the middle of the screen for 1250 ms. The word was replaced by a fixation point and another equation was presented. This procedure continued until the set size was completed. Sets consisted of two to six equation/word pairs. Participants started with sets of two pairs, and the set size increased after every three sets. After the completion of a set, the word RECALL appeared in the center of the screen. At this point, participants had to type into the computer all the words that they could recall from the set. The order of their response did not matter, but they were instructed not to enter the last word of the set first. When they finished recalling as many words as possible, they pressed the ESC key to enter their responses and begin the next set.

3.4.2.2 Data Analysis

Mean RTs were calculated separately for trials in which participants had to respond "yes" and "no". Outliers, identified as responses 2.5 *SD* above or below the mean, were excluded from analysis. The O-span score for each individual was calculated by taking the total number of correctly recalled words for trials on which the mathematical equation was correctly judged.

Together with the measures of proficiency, the measures of individual cognitive resource differences provided a more complete picture of the subset of bilinguals tested in the current study. These variables were then used as predictors in regressions to better evaluate the impact of proficiency and cognitive resources on L2 gender processing.

Chapter 4

Experiment 1: Gender Processing in Simple Nouns

4.1 Experiment 1A: English-German L2 Learners

The first experiment set out to investigate the sensitivity of L2 learners of German to grammatical gender. As discussed in Chapters 1 and 2, past research on L2 gender processing has shown that L2 learners display a particular difficulty with grammatical gender. Previous studies on native English speakers learning a gendered L2 have, by and large, used off-line measures (e.g., Sabourin, Stowe, & de Haan, 2006), more complex sentence structures (e.g., Taraban & Kempe, 1999), or have focused on gender systems with more transparent gender assignment (e.g., Tokowicz & MacWhinney, 2005). Some have argued for a dissociation in L2 learners between being able to identify the gender of a noun and being able to use it in context (e.g., Guillelmon & Grosjean, 2001; Sabourin et al., 2006) or have claimed that while L2 learners may learn compensatory strategies to perform at native-like levels, they will never attain native-like processing (Carroll, 1989; Ellis, 1984; Hawkins & Franceschina, 2004). By using a translation recognition task, the present study hoped to provide an on-line measure of gender sensitivity in a task that should be relatively easy for second language learners at different levels of proficiency to complete. Furthermore, although the translation recognition task requires judgments of word pairs out of sentence context, it has been shown to be sensitive to grammatical class (Sunderman & Kroll, 2006).

A second goal of Experiment 1A was to evaluate the role of proficiency in modulating sensitivity to gender. With increasing proficiency, participants should display not only faster and more accurate processing of translation equivalents, but also increasing sensitivity to the matched and mismatched gender conditions. In particular, participants with higher proficiency should show longer response latencies in rejecting translations with the right article but wrong noun since the initial presentation of the gender should confirm the anticipated translation only to be violated with the presentation of the noun. This violation between expectation (and being ready to make a “yes” response) and presentation (and having to then respond with a “no” response) should result in processing costs as participants revise their response. Responses in this condition should contrast strongly with response times to wrong article, wrong noun translations that require no response revision. In contrast, lower proficiency participants, who may not be as sensitive to grammatical gender, should show no difference in processing these two conditions since the defining difference between the conditions is the match or mismatch of gender to the correct translation. Finally, participants with lower L2 proficiency should show particular difficulty rejecting items with the wrong article and the correct noun translations assuming they are not sensitive to gender. This could result in lower accuracy scores on this condition due to false alarms (i.e., responding “yes” even though the translation is wrong) as well as slower response times. However, if a maturational (i.e., critical period) account of L2 learning is correct, then participants, regardless of proficiency, should not show gender sensitivity in these conditions since all of the English-German participants in this group are late learners of German.

4.1.1 Method

4.1.1.1 Participants

Eighty-two native and dominant English speakers with intermediate to advanced proficiency in German participated in Experiment 1A and were recruited from several large universities in the north-east and south-west of the United States. All 82 were included in the analyses.

4.1.1.2 Materials

Sixty critical simple English nouns were included. Each noun was paired with three German translations: A. an incorrect translation item which matched the correct translation in gender (e.g., for the English 'the rabbit' whose true translation is 'der$_{masculine}$ Hase', THE RABBIT – DER$_{masculine}$ KNOPF$_{button}$), B. an incorrect translation item which did not match the correct translation in gender (e.g., THE RABBIT – DAS$_{neuter}$ BRETT$_{board}$) and C. an item which was the correct translation but was assigned an incorrect gender (e.g., THE RABBIT – *DAS$_{neuter}$ HASE). The three possible German translations for each English word were matched as closely as possible on frequency and word length. Care was taken that the incorrect gender assignment could not form a plausible gender assignment in a different case (e.g., zero plurals or genitive case or dative case).

The English translations were divided into matching triads which were item-matched on gender, frequency and word length (e.g., THE RABBIT (der Hase) – THE

SCARF (der Schal) – THE KNIGHT (der Ritter)), and the same incorrect translations were assigned for conditions A and B for these three items (e.g., DER KNOPF (the button), DAS BRETT (the board)). The items were then cycled through in such a way so that across three participants, all items and all conditions would be seen, and each participant would see all the English words, but would only see a given incorrect German translation form once. However, by virtue of the triads across participants sharing two of the three German incorrect translation forms, each participant would also see almost all the same German incorrect translation forms, allowing us to control for lexical familiarity (see Table 4.1 below for an example of this distribution).

Table 4.1: Example distribution of items across three conditions for the critical English words SCARF (der Schal), RABBIT (der Hase), and KNIGHT (der Ritter). DER KNOPF (the button) and DAS BRETT (the board) are each grammatical in German on their own.

	Condition		
	1 (German is grammatical) right article, wrong noun (CI)	2 (German is ungrammatical) wrong article, right noun (IC)	3 (German is grammatical) wrong article, wrong noun (II)
Participant 1:	THE SCARF – DER KNOPF	THE KNIGHT – DAS RITTER	THE RABBIT – DAS BRETT
Participant 2:	THE RABBIT – DER KNOPF	THE SCARF – DAS SCHAL	THE KNIGHT – DAS BRETT
Participant 3:	THE KNIGHT – DER KNOPF	THE RABBIT – DAS HASE	THE SCARF – DAS BRETT

Note. In Condition 2, the presented German article and noun are ungrammatical, whereas in Conditions 1 and 3, the German article and noun pair are grammatical, although they are the wrong translation of the English word.

T-tests were performed to ensure that across conditions and participant lists as well as across conditions within a given participant list, German frequency and word length as well as English frequency, age of acquisition, word length, and imageability were not significantly different. Table **4.2** below illustrates the mean lexical properties of the English items. Because the items were cycled through the conditions of the experiment, they were the same across conditions. A full listing of the items is given in Appendix D.

Items were also avoided which would be considered false cognates, although a subset of 15 cognate items were included. 60 filler English simple nouns (45 noncognate and 15 cognate) with correct translations ("yes" items) were selected which matched in frequency to the critical "no" items across lists.

Table 4.2: Mean Lexical Properties for the critical English words

Variable	Mean	*SD*
Frequency (per million words)[a]	105.8	191.1
Log Frequency[a]	1.7	0.5
Frequency (per million words)[b]	104.8	162.7
Familiarity (100-700)[c]	551	48.9
Length in syllables	1.5	0.8
Length in characters	5.3	1.7
Age of Acquistion (100-700)[d]	296.5	98.4
Imageability (100-700)[c]	510.8	102.8

Note. [a]Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995); [b]Kučera, H., & Francis, W. N. (1967); [c]Coltheart (1981); [d]Gilhooly, K. J. & Logie, R H. (1980)

The distribution of gender in the critical items corresponded to the natural distribution of gender in German where 50% of all words are masculine, 30% feminine, and 20% neuter. However, overall, including fillers, participants saw close to an equal distribution of gender across lists. The reason that the fillers were designed this way was to avoid a strategic bias based on the presentation of the article prior to the German translation.

In addition, 112 compound nouns were added as fillers to be used for another experiment (see Experiment 3) and will not be discussed further in this chapter. None of the individual constituents of the compounds in the filler items appeared as simple nouns. In this way, participants only saw a given constituent once in the entire experiment.

4.1.1.3 Procedure

Participants were tested individually in a quiet room on a PC using E-prime stimulus presentation software (Schneider, Eschman, & Zuccolotto, 2002). For the translation recognition task, text appeared in white in bold Courier New 24 font size on a black background. Prior to each trial, a fixation sign (+) appeared at the center of the computer screen. When participants were ready to begin, they pressed the spacebar to initiate the trial. First an article and noun in English appeared in the center of the screen, presented word by word, followed by an article and noun in German. The English article remained on the screen for 200 ms followed by an inter-stimulus interval (ISI) of 100 ms and then the English noun for 700 ms and an ISI of 700 ms. The German article then appeared for 200 ms, with an ISI for 100 ms, and then the German noun appeared for 700 ms. After presentation of the noun, participants were asked to indicate as quickly as possible whether the German noun phrase was the correct translation of the English noun phrase by pressing a "yes" button or a "no" button on the keyboard. They were given 4000 ms from onset of the German noun to respond before the next fixation sign appeared (see Figure 4.1 for an illustration of a trial). Participants completed 15 practice items including all three types of incorrect translation types. Before beginning the actual experiment, they were verbally reminded to pay attention to the accuracy of the German article paired with the German noun, and to reject the translation as incorrect if the article did not agree with the noun.

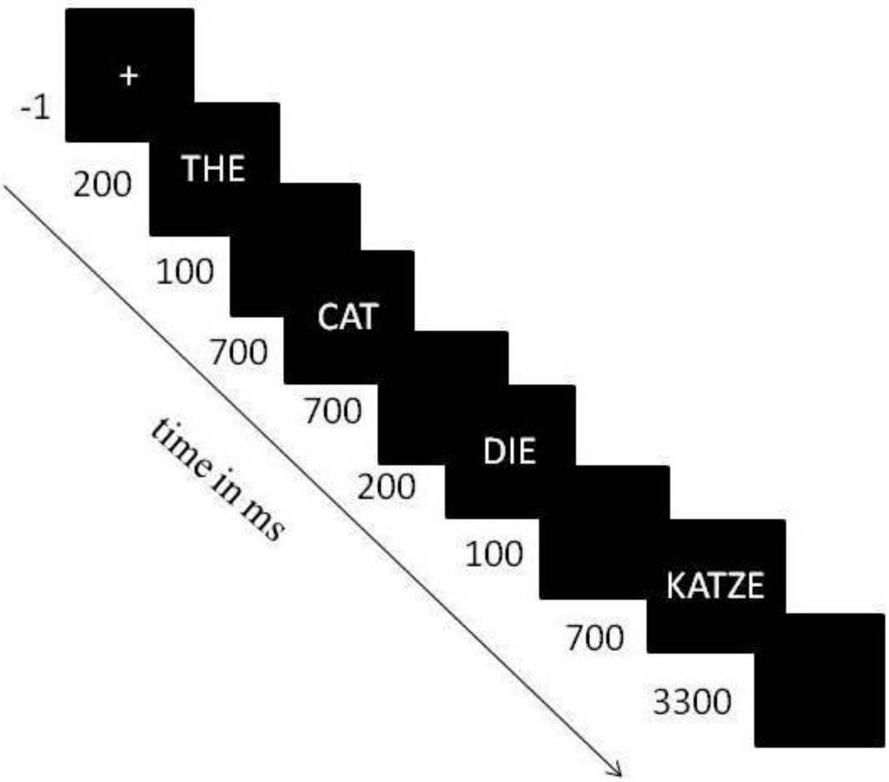


Figure 4.1: An illustration of a trial in the simple noun experiment.

After completion of the translation recognition task, participants completed the Operation Span task followed by the Gender Assignment task. They then completed the Simon task and a Language History Questionnaire.

4.1.2 Results and Discussion

Two types of analyses were conducted. First, general analyses over the entire data were run. In a second selective set of analyses, for each participant, only items identified with the correct gender during the later gender assignment task were included in the translation recognition analyses. Since the pattern of results was identical in both analyses, only the overall analyses are reported here. Whenever the assumption of sphericity was violated, as indicated by Mauchly's test, the degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity. Incorrect responses (18.5 % of the data), as well as items with response latencies below 300 ms and above 3000 ms or deviating two and a half *SD*s from a participant's mean (5.4% of the data) were excluded from reaction time analyses.

Reaction times as well percent accuracy scores by participant (*F*1) and item (*F*2) were entered into a one-way repeated measures analysis of variance (ANOVA) with simple noun translation type (correct gender, incorrect noun; incorrect gender, correct noun; incorrect gender, incorrect noun) as the independent variable.

4.1.2.1 Latencies

The results revealed a main effect of translation type ($F1(1.29, 102.23) = 70.84$, $MSE = 28088.43$, $p < .001$; $F2(1.56, 91.73) = 65.77$, $MSE = 25513.77$, $p < .001$). Pairwise comparisons using a Bonferroni adjustment indicated that participants rejected right article, wrong noun translations faster than wrong article, right noun translations ($p < .001$) and also rejected wrong article, wrong noun translations faster than wrong article, right noun translations ($p < .001$). The difference in reaction time between wrong article, right noun and wrong article, wrong noun translations approached significance ($p < .1$). Pairwise comparisons by items showed a similar pattern of results, with faster reaction times for right article, wrong noun translations than wrong article, right noun translations and significantly faster reaction times for wrong article, wrong nouns than wrong article, right nouns ($p < .001$). There was no significant difference between right article, wrong noun translations and wrong article, wrong noun translations ($p > .1$). Mean participant latencies are displayed in Figure 4.2.

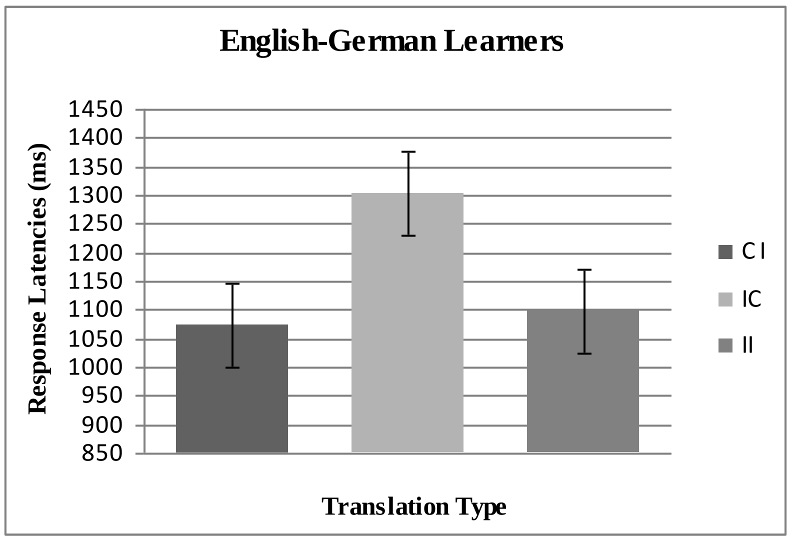


Figure 4.2: Mean translation latencies for English-German learners as a function of translation condition where CI = right article, wrong noun; IC = wrong article, right noun; and II = wrong article, wrong noun.

4.1.2.2 Accuracy

The results for accuracy data revealed corroborating results to those found for the latencies, with a main effect of translation type for both participant and item data ($F1(1.34, 108.83) = 454.32$, $MSE = 237.43$, $p < .001$; $F2(1.48,87.49) = 499.60$, $MSE = 144.91$, $p < .001$). Pairwise comparisons for participant data using a Bonferroni adjustment showed that both the right article, wrong noun translation condition and the wrong article, wrong noun translation condition were significantly more accurate than the wrong article, right noun condition (*p*s < .001) but were not significantly different from each other (*p* > .1). The same pattern of data was found for item data. Mean participant accuracy scores are displayed in Figure 4.3.

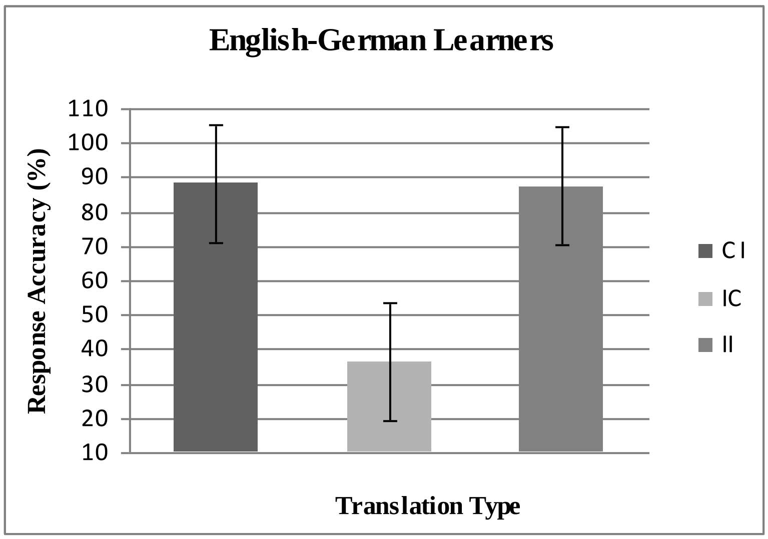


Figure 4.3: Mean translation accuracy for English-German learners as a function of translation condition where CI = right article, wrong noun; IC = wrong article, right noun; and II = wrong article, wrong noun.

Overall, these results suggest that these participants were not sensitive to grammatical gender processing in the translation recognition task. Latencies and accuracy scores did not significantly differ between the right article, wrong noun translation condition and the wrong article, wrong noun translation condition, again suggesting that these participants were not sensitive to gender across translation. Furthermore, the particular difficulty participants faced in rejecting the wrong article, right noun translations, as reflected in increased latencies and decreased accuracy scores, indicates that if anything, they were particularly tuned to the semantics of the translation task and not able to focus in on the gender component of the task. It is possible, however, that a subset of participants in this group with higher proficiency might begin to show an increased ability to process grammatical gender.

To investigate the role of proficiency in gender processing, participants were grouped according to proficiency using a median split based on their percent accuracy score on the gender assignment task. Although not completely independent of the translation recognition task since the critical German items were the same, the gender assignment task was used as it represented the most accurate indication of grammatical, and particularly gender, knowledge in this population of L2 German learners. Table 4.3 shows mean scores between the two proficiency groups on various aspects of their language history questionnaire, as well as on cognitive processing as measured by the Simon task and the operation-span task.

Table 4.3: Characteristics of English-German participants in Experiment 1A by Proficiency Group

	Less proficient English-German (n = 42)	More proficient English-German (n = 40)	Significance of group difference
Age (years)	28.2	27.0	*ns*
	(12.4)	(9.8)	
L1 self-ratings (10-pt scale)	9.5	9.6	*ns*
	(0.7)	(0.5)	
L2 self-ratings (10-pt scale)	6.6	7.8	0.001
	(1.7)	(1.0)	
L2 age of acquisition (years)	15.4	14.5	*ns*
	(7.8)	(5.0)	
Months of adult L2 immersion	13.0	17.7	*ns*
	(20.3)	(30.9)	
Simon effect (ms)	27.7	43.2	0.020
	(35.0)	(22.3)	
Operation span (1-60)	48.0	46.9	*ns*
	(7.6)	(7.6)	
Gender assignment (ms)	1646.0	1293.0	0.001
	(395.5)	(203.2)	
Gender assignment accuracy (%)	50.6	70.9	0.001
	(6.6)	(7.4)	

Note. Standard Deviations are in parentheses

Reaction times on the translation task, as well percent accuracy scores, were entered into a 3 x 2 mixed ANOVA with simple noun translation type (correct gender, incorrect noun; incorrect gender, correct noun; incorrect gender, incorrect noun) as a within-participants variable and proficiency group (high vs. low) as a between-participants variable. Analyses were conducted treating both participants and items as random variables. In the item analysis, proficiency group was a within-item variable. Item analyses were based on the English noun, for which a German translation was presented in each of the three critical conditions over a full counterbalance (see description of materials above). As a result, translation type remained a within-item variable as in the participant analyses.

4.1.2.3 Latencies by Proficiency

Levene's test of homogeneity of variance indicated that the variances between proficiency groups for participant (*F*1) data were significantly different for the right article, wrong noun condition ($F(1,78) = 7.16$, $p < .01$) as well as the wrong article, wrong noun condition ($F(1,78) = 9.35$, $p < .01$) and approached significance for the wrong article, right noun condition ($F(1,78) = 2.81$, $p < .1$). The data were thus corrected by taking the reciprocal square root of each value ($1/\sqrt{X_i}$) in each condition. Results revealed a main effect of translation type ($F1(1.35, 105.58) = 58.57$, $MSE = 3.89E-006$, $p < .001$; $F2(1.56, 84.09) = 81.78$, $MSE = 47910.51$, $p < .001$) and a main effect of proficiency group ($F1(1,78) = 6.24$, $MSE = 4.58E-005$, $p < .05$; $F2(1, 54) = 82.28$, $MSE = 29135.18$, $p < .001$). There was no significant interaction between translation type and group. Pairwise comparisons of the participant data of the types of translations using a Bonferroni correction revealed that the right article, wrong noun condition and the wrong article, wrong noun condition were significantly different from the wrong article, right noun condition (*p*s <.001) while their difference from each other approached significance ($p < .1$). Pairwise comparisons of the item data yielded identical results. Uncorrected participant means by proficiency group are reported in Figure 4.4.

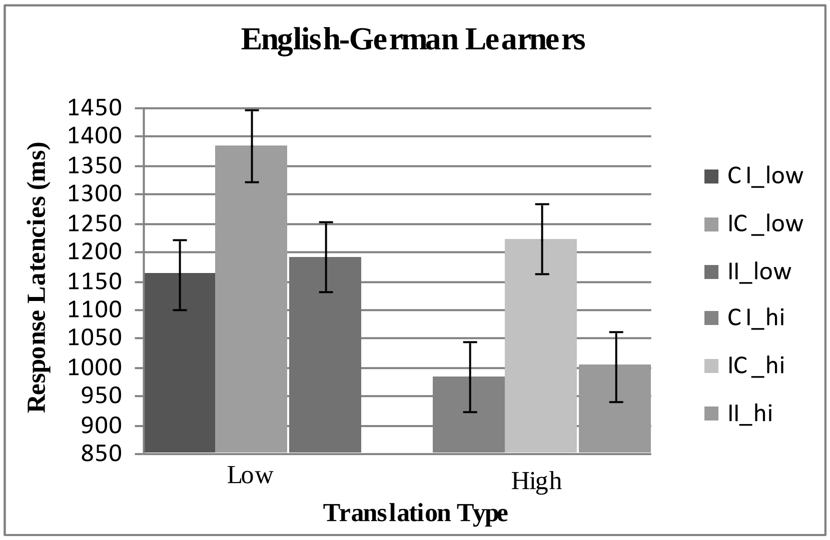


Figure 4.4: Mean translation latencies for English-German learners as a function of translation condition and proficiency where CI = right article, wrong noun; IC = wrong article, right noun; II = wrong article, wrong noun; hi = high proficiency; and low = low proficiency.

4.1.2.4 Accuracy by Proficiency

Levene's test of homogeneity of variance indicated that the accuracy variances between proficiency groups for participant (*F*1) data were significantly different for all three translation conditions (right article, wrong noun condition: $F(1,80) = 17.64$, $p < .001$; wrong article, wrong noun condition: $F(1,80) = 11.89$, $p < .01$; wrong article, right noun condition: $F(1,80) = 4.09$, $p < .05$). The participant data were corrected by taking the arc sine transformation ($\arcsin\sqrt{(X_i/100)}$) of each value in each condition. Results revealed a main effect of translation type ($F1(1.65, 131.69) = 370.75$, $MSE = 124.02$, $p < .001$; $F2(1.55, 91.64) = 491.46$, $MSE = 275.32$, $p < .001$) and a main effect of proficiency group ($F1(1.65, 80) = 43.65$, $MSE = 197.13$, $p < .001$; $F2(1, 59) = 91.13$, $MSE = 178.68$, *p* < .001) but no interaction. Pairwise comparisons with a Bonferroni correction showed that participants were significantly more accurate in the right article, wrong noun and the wrong article, wrong noun conditions than the wrong article, right noun condition (both *p* < .001), but that they were not significantly different from each other (*p* > .1). Item comparisons yielded identical levels of significance. Uncorrected participant means are reported in Figure 4.5**.**

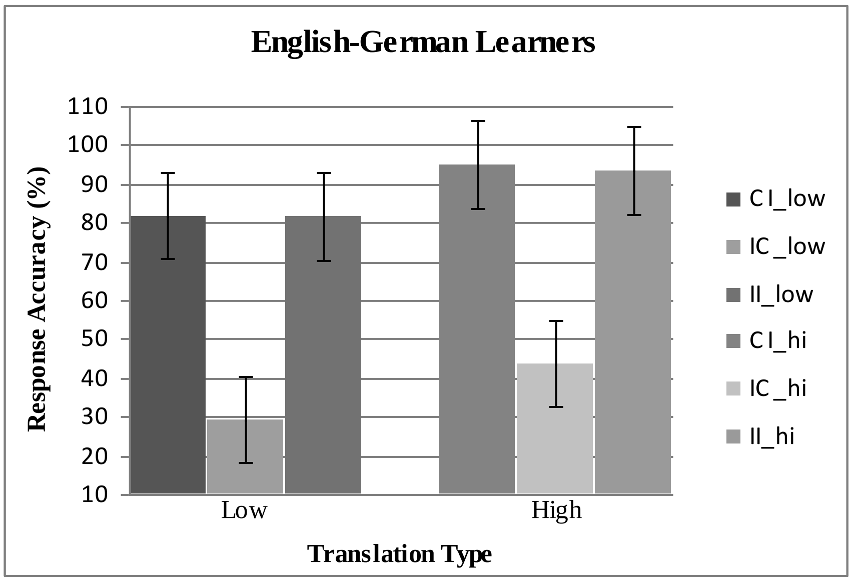


Figure 4.5: Mean translation accuracy for English-German learners as a function of translation condition and proficiency where CI = right article, wrong noun; IC = wrong article, right noun; II = wrong article, wrong noun; hi = high proficiency; and low = low proficiency.

The results of these analyses showed that while participants increased in both speed and accuracy in translation recognition with increased proficiency, their sensitivity to the finer nuances of gender did not change as a function of proficiency. Regardless of the level of proficiency, the right article, wrong noun condition and the wrong article, wrong noun condition did not differ significantly from each other. If anything, mean trends suggest that latencies were shorter in the right article, wrong noun condition than the wrong article, wrong noun condition, against the predicted direction of results. Differences between these two conditions and the wrong article, right noun condition appeared to increase with increasing proficiency, at least in the item analysis, possibly indicating increased conflict between choosing the appropriate semantics and focusing in on the task of gender processing.

4.2 Experiment 1B: German-English L2 Learners

The adaptation of the translation recognition task used in Experiment 1A has not been used in past research with native speakers of German learning English. The goal of the next experiment was to determine whether native speakers of German would be sensitive to the matching or mismatching gender conditions in the tranlation recognition task. Because the order of presentation was kept constant over the two experiments (English first, German second), the direction of translation changed for native German participants so that they were translating from the L2 into their dominant language. The rationale behind this decision was to allow both participant groups to anticipate the German translation after having seen the English probe. Due to the difference in translation direction, a direct comparison between language groups cannot be made in the present study using the present translation recognition paradigm. However, results of the native German speakers should give some indication whether conditions such as the comparison of the right article, wrong noun condition and the wrong article, wrong noun condition could be sensitive to gender knowledge and gender processing. Results of this study can therefore help address whether null results between these two conditions in

Experiment 1A were due to the language group and/or proficiency, or due to task limitations.

4.2.1 Method

4.2.1.1 Participants

Twenty-eight native German speakers who were L2 learners of English were recruited for the experiment from a large university in Germany. Seven native German L2 learners of English were recruited at a large university in the northeast of the United States and were included in the sample even though they were immersed in the L2 environment, as their data did not pattern differently from the other sample. A total of 35 participants were therefore included for analysis.

4.2.1.2 Materials

The same materials as in Experiment 1A were used.

4.2.1.3 Procedure

The same procedure was used as in Experiment 1A.

4.2.2 Results and Discussion

Reaction times as well percent accuracy scores by participant (*F*1) and item (*F*2) were entered into a one-way repeated measures analysis of variance (ANOVA) with simple noun translation type (correct gender, incorrect noun; incorrect gender, correct noun; incorrect gender, incorrect noun) as the independent variable. Whenever the assumption of sphericity was violated, as indicated by Mauchly's test, the degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity. Incorrect responses, as well as items with response latencies below 300 ms and above 3000 ms or deviating 2.5 *SD*s from a participant's mean, were excluded from reaction time analyses.

4.2.2.1 Latencies

The results revealed a main effect of translation type ($F1(2, 68) = 35.98$, $MSE = 2888.07$, $p < .001$; $F2(2, 118) = 8.46$, $MSE = 19284.23$, $p < .001$). Pairwise comparisons using a Bonferroni adjustment indicated significant differences between all conditions (*p*s $< .05$) where participants rejected wrong article, right noun translations fastest, followed by wrong article, wrong noun translations. Participants were slowest to reject items in the right article, wrong noun condition, suggesting sensitivity to grammatical gender across translations. Pairwise comparisons by items showed a similar pattern of results, with significant differences between the right article, wrong noun condition and the wrong article, right noun condition ($p < .01$) with faster latencies in the latter condition. Items in the wrong article, right noun condition were also rejected faster than items in the wrong article, wrong noun condition ($p < .05$). Mean participant latencies are displayed in Figure 4.6.

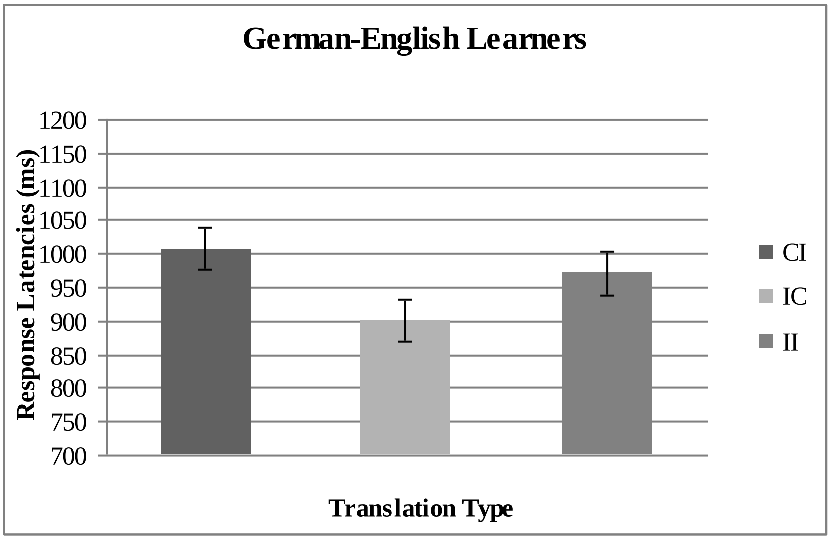


Figure 4.6: Mean translation latencies for German-English learners as a function of translation condition where CI = right article, wrong noun; IC = wrong article, right noun; and II = wrong article, wrong noun.

4.2.2.2 Accuracy

Because accuracy data were at ceiling for this group of participants, scores were first normalized by taking the arc sine transformation ($\arcsin\sqrt{(X_i/100)}$) of each value in each condition for both participant and item data. The results for accuracy data revealed no significant differences between conditions ($F1 < 1$; $F2(2, 118) = 1.48$, $MSE = 104.93$, $p > .1$), although there was a trend in the means by participant and item for slightly lower accuracy scores in the right article, wrong noun condition, corroborating findings in the latencies. Uncorrected mean participant accuracy scores are displayed in Figure 4.7.

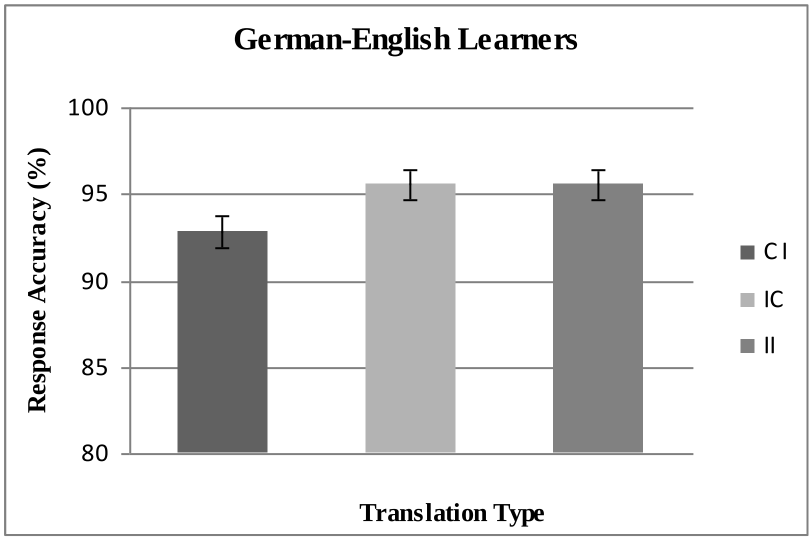


Figure 4.7: Mean translation accuracy for German-English learners as a function of translation condition where CI = right article, wrong noun; IC = wrong article, right noun; and II = wrong article, wrong noun.

The results of these initial analyses suggest that native German speakers who are learners of English are sensitive to grammatical gender across translations. Particularly the difference between the right article, wrong noun condition and the wrong article, wrong noun condition provides compelling evidence for this sensitivity. Results seem analogous to Stroop-like effects, such that participants in this situation must override prepotent responses to accept the translation as correct based on the validity of the gender after the appearance of the noun indicates a translation violation. The fact that response times are faster in the wrong article, right noun condition than the wrong article, wrong noun condition could be due to additional semantic priming in the former condition.

To explore the potential impact of proficiency level on task performance, as was done in Experiment 1A, the German-English bilinguals were also split up into high and low proficiency groups.

As the gender assignment task did not provide a measure of their L2 grammatical proficiency, groupings were based on whether participants had studied English at the university level or not. Table 4.4 shows mean scores between the two proficiency groups on various aspects of their language history questionnaire, as well as on cognitive processing as measured by the Simon task and the Operation-Span task.

Table 4.4: Characteristics of German-English participants in Experiment 1A by Proficiency Group

	Less proficient German-English (n=15)	More proficient German-English (n=20)	Significance of group difference
Age (years)	26.7	28.4	*ns*
	(2.6)	(7.4)	
L1 self-ratings (10-pt scale)	9.2	9.6	.090
	(0.7)	(0.7)	
L2 self-ratings (10-pt scale)	6.7	8.1	.002
	(1.4)	(1.0)	
L2 age of acquisition (years)	10.3	10.5	*ns*
	(1.9)	(1.4)	
Months of adult L2 immersion	5.6	7.8	*ns*
	(5.0)	(7.1)	
Simon effect (ms)	34.1	44.2	*ns*
	(21.7)	(19.9)	
Operation span (1-60)	46.4	46.9	*ns*
	(7.0)	(7.3)	
Gender assignment (ms)	918.0	961.0	*ns*
	(81.2)	(106.3)	
Gender assignment accuracy (%)	96.4	96.7	*ns*
	(1.2)	(1.5)	

Note. Standard Deviations are in parentheses

Reaction times on the translation task as well percent accuracy scores by participant ($F1$) and item ($F2$) were entered into a 3 x 2 mixed ANOVA with simple noun translation type (correct gender, incorrect noun; incorrect gender, correct noun; incorrect gender, incorrect noun) as a within-participants variable and proficiency group (high vs.

low) as a between-participants variable. In the item analysis, proficiency group was a within-item variable. Item analyses were based on the English noun, for which a German translation was presented in each of the three critical conditions over a full counterbalance (see description of materials in Experiment 1A above). As a result, translation type remained a within-item variable as in the participant analyses.

4.2.2.3 Latencies by Proficiency

Results revealed a main effect of translation type ($F1(2, 66) = 34.24$, $MSE = 2974.25$, $p < .001$; $F2(2, 118) = 9.23$, $MSE = 39084.21$, $p < .001$), a main effect of proficiency group only in the item analysis ($F1 < 1$; $F2(1, 59) = 120.42$, $MSE = 9939.79$, $p < .001$) and no significant interaction between translation type and group (*F*s < 1). As there was no group x condition interaction, pairwise comparisons of the participant and item data of the types of translations using a Bonferroni correction were identical to results from ungrouped data with significant differences between all three conditions (*p*s <.05). Pairwise comparisons of the item data of the groups revealed significantly faster reaction times in the higher proficiency group, against the mean trends of participant data ($p < .001$). Participant means by proficiency group are reported in Figure 4.8.

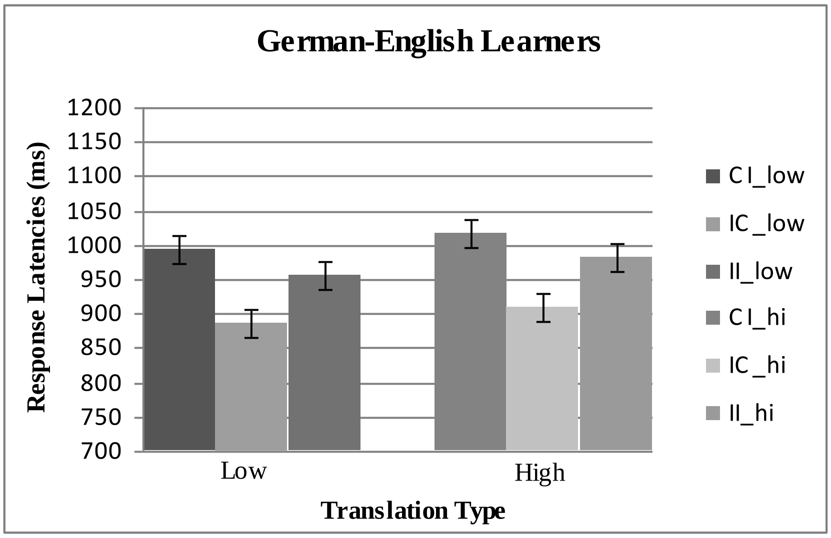


Figure 4.8: Mean translation latencies for German-English learners as a function of translation condition and proficiency where CI = right article, wrong noun; IC = wrong article, right noun; II = wrong article, wrong noun; hi = high proficiency; and low = low proficiency.

4.2.2.4 Accuracy by Proficiency

As in the ungrouped data, the participant data were corrected by taking the arc sine transformation ($\arcsin\sqrt{(X_i/100)}$) of each value in each condition. Results revealed no main effect of translation type and no effect of proficiency group (Fs < 1). Mean trends again corroborated finding in the latencies, with lower accuracy in the right article, wrong noun condition than in the other two conditions. There was, however, a significant interaction between group and proficiency in the item analysis ($F1 < 1$; $F2(2, 118) = 3.46$, $MSE = 151.44$, $p < .05$). To follow up the significant interaction, separate ANOVAs were run for each proficiency group. Results showed a significant effect of condition only for the lower proficiency group (low: $F2(2, 118) = 4.08$, $MSE = 167.20$, $p < .05$; high:

$F2(2, 118) = 1.08$, $MSE = 135.53$, $p > .1$). Pairwise comparisons with a Bonferroni correction revealed significantly lower accuracy scores in the right article, wrong noun condition than the wrong article, right noun condition ($p < .05$). No other differences were significant. Uncorrected participant means are reported in Figure 4.9.

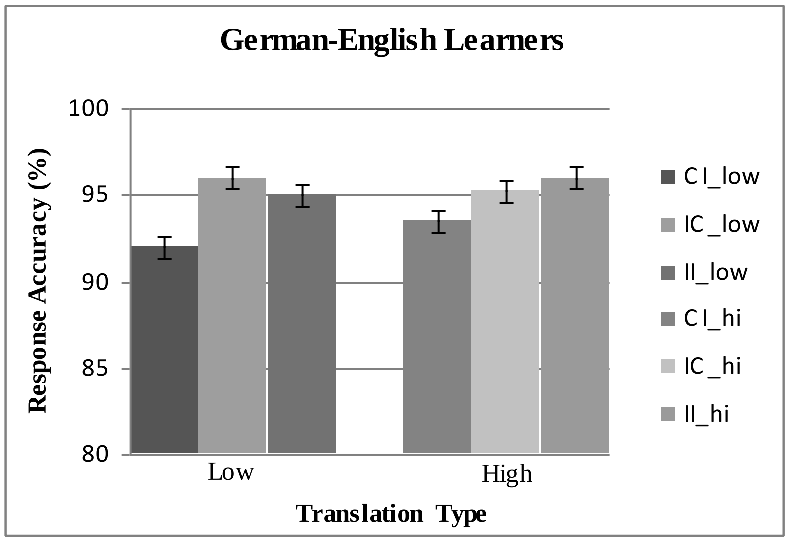


Figure 4.9: Mean translation accuracy for German-English learners as a function of translation condition and proficiency where CI = right article, wrong noun; IC = wrong article, right noun; II = wrong article, wrong noun; hi = high proficiency; and low = low proficiency.

The results of the proficiency analyses suggest that in native speakers of German, there appears to be little consequence of additional proficiency in their L2 English. Note that the range of proficiency in this group is much smaller than for the English-German group, and their L2 proficiency overall higher. Overall, participants in the higher proficiency group were faster to respond than counterparts in the lower proficiency group, indicating, not surprisingly, that processing speed decreased as a function of L2 proficiency. While there was a hint of an interaction of condition with proficiency group in the item analysis of the accuracy data, the pattern of the result was in the direction of previous results, showing a decreased accuracy for items in the right article, wrong noun condition compared to the other conditions. Results in the analysis by proficiency corroborated previous analyses indicating that the right article, wrong noun condition posed particular difficulties for native German speakers both in processing speed and accuracy. These results suggest that this group of bilinguals was sensitive to gender across translations.

Overall, Experiment 1B provides evidence that the translation recognition task is sensitive to gender processing, at least in native speakers of German. Similar to results with native English speakers in Experiment 1A, participants in Experiment 1B showed an overall decrease in response latencies with increasing L2 proficiency, but no interaction of proficiency with translation conditions. Significantly, overall results and results by group demonstrated increased latencies for the right article, wrong noun condition over the wrong article, wrong noun condition, suggesting that the German-English participants were revising their response after the presentation of the German noun.

4.3 General Discussion

The main purpose of the present experiment was two-fold. Primarily, the experiment sought to characterize the type of gender processing of L2 learners of German whose native language, English, does not have grammatical gender. Secondly, the experiment investigated the role of proficiency in modulating sensitivity to gender in late L2 learners. The results of Experiment 1A suggest that, overall, this group of L2 learners was not sensitive to gender in the current paradigm. Participants showed no significant difference in latencies or accuracy scores between the right article, wrong noun condition and the wrong article, right noun condition. If anything, mean trends suggest a pattern of results in the opposite direction, with faster reaction times in the right article, wrong noun condition over the other conditions. Participants also demonstrated particular difficulty with the wrong article, right noun condition as reflected in significantly slower response latencies and lower accuracy scores which indicate high false alarm rates in this condition. Similar to previous results in translation recognition, these results suggest that even learners are sensitive to semantics in on-line L2 language processing (e.g., Sunderman & Kroll, 2006). Anecdotally, several participants upon debriefing indicated that they had realized an error in their choice in this condition milliseconds after having made a response. Consistent with other research and L2 theories of language processing, this observation points to late effects of gender processing which are more open to strategies. With a high level of L2 proficiency, learners may eventually acquire a strategy for computing gender assignment but that strategy may not be available during the initial stages of identifying the meaning of words in the L2.

Experiment 2A, with more proficient German speakers of English, demonstrated that the translation-recognition task itself is sensitive to gender processing. While no strong comparison between language groups can be made, the results leave open the possibility that participants with a more native-like command of German might begin to show sensitivity to gender using this task. A post-hoc investigation of data patterns in the means of the English-German learners revealed that 31 out of the 82 participants indeed showed longer latencies in the right article, wrong noun condition than the wrong article, wrong noun condition, and 18 showed differences above the median difference in native

German speakers. These findings agree with other estimates of native-like L2 speakers in a sample of L2 learners which range from 5 to 20 % of the sample (as quote by Birdsong & Mollis, 2001). As Long (1990) points out, even the presence of a single native-like L2 speaker would refute the critical period and suggest that native-like attainment is possible, perhaps not for all, but at least for some.

A possibility to be explored is whether L2 German learners, although they show little sensitivity to on-line grammatical gender processing in the current experimental paradigm, might nevertheless be able to use grammatical gender in a task more under their explicit control. Previous research in both monolinguals and bilinguals has suggested that speakers of a language with grammatical gender are sensitive to regularities, such as the phonological features of a noun in assigning gender (Bordag, et al., 2006; Bordag & Pechmann, 2007; Schiller et al., 2003; Schwichtenberg & Schiller, 2004). Interestingly, the study by Schiller and colleagues (2003) found a dissociation between behavioral and ERP results, where native speakers of German completing a gender assignment task showed a sensitivity to phonological rules behaviorally, but not in ERP measures. They did, however, find a strong effect of semantic gender information on gender decision (see also Hofmann, 2005, for similar results and similar dissociations). These results suggest that probabilistic rules play an important role in determining syntactic gender decisions, particularly at later processing stages, then do not appear to influence initial gender processing. The question then remains how native speakers of a language without gender might use phonological rules in order to learn gender in an L2 whose gender assignment appears to be, at least superficially, relatively arbitrary. This is the question to be addressed in Chapter 5.

Chapter 5

Analysis of Gender Assignment Data from Experiment 1

The results of Experiment 1A suggest that while performance in rejecting nouns with incorrect gender is only barely above chance, English-German participants are still able to recognize incorrect gender assignment some of the time. What mechanisms might help learners boot-strap their way into the gender system? One possibility is that phonological features might provide particularly salient cues to the L2 learner in deciding the gender of a given word. For instance, in German, the *–e* ending on a noun typically indicates feminine gender. This cue is the first gender cue that children learning German acquire (MacWhinney, 1978; Mills, 1986) and is also the most frequent phonological gender cue in German, whereby 90% of words ending in *–e* are feminine. Previous research by Bordag and colleagues with bilingual L2 speakers of gendered languages indicates that bilinguals are, in fact, sensitive to phonological cues (English-German: Bordag et al., 2006; Czech-German: Bordag & Pechmann, 2007). In Bordag et al. (2006), results of a gender picture-naming task and a grammaticality judgment task with L1 German speakers showed no sensitivity to these phonological gender rules, while intermediate speakers of German (L1 English) revealed an influence of the phonological noun form in both reaction time and accuracy. Other monolingual German research, such as Hohlfeld (2006), suggests that this phonological information only comes into play at later processing stages such as off-line tasks, but is not used on-line, similarly suggesting different processing strategies for monolingual and bilingual speakers, or at least the possibility that learners are on a different point of a learning continuum than monolingual speakers (see Bordag et al., 2006, and Taraban and Kempe, 1999, for similar views). To further investigate the extent of gender knowledge in this sample of bilinguals, gender knowledge of L2 learners was assessed in a metanlinguistic task This task was similar to the sort used in past research that has been used to determine whether apparent failures to use subtle aspect of the grammar in the L2 reflect a failure of knowledge or a failure of processing. The same sample of participants as in Experiments 1A and 1B were tested, allowing a similar comparison to the one made by Bordag and colleagues.

5.1 Gender Assignment: English-German learners

5.1.1 Method

5.1.1.1 Participants

The data from Experiment 1A served as the basis of the current analysis, so that data from the same 82 English-German learners in Experiment 1A was used.

5.1.1.2 Materials

Even though participants assigned gender to all critical items in Experiment 1, seventy-five of the items were selected for further analysis according to the following criteria. Using the phonological gender categorization of nouns in Bordag et al. (2006), nouns from the gender assignment task were divided into three categories: typical, ambiguous, and atypical. While most feminine nouns end in shwa (-*e*) and most masculine and neuter nouns end in a consonant, there are a few exceptions to this distribution, with feminine nouns ending in a consonant or masculine and neuter nouns ending in –*e*. Based on this distribution, nouns in the typical category were feminine nouns ending in –*e*. Nouns in the ambiguous category were masculine or neuter nouns ending in a consonant. Nouns categorized as atypical were feminine nouns ending in a consonant, or masculine or neuter nouns ending in –*e*. Twenty-five nouns were selected from each category and were matched on German word length (as well as English word length of the translation), Familiarity, English Age of Acquisition, German and English frequency, and Imageability (see Table 5.1). The full set of items are available in Appendix E.

Table 5.1: Characteristics of the 75 items in the Gender Assignment task by gender category

Variable	Typical (1)	Ambiguous (2)	Atypical (3)	Sig. 1 vs. 2	Sig. 2 vs. 3	Sig. 1 vs. 3
German frequency[a]	10.8 (2.2)	10.8 (1.8)	9.9 (2.5)	0.89	0.16	0.16
German length	6.0 (1.5)	5.9 (1.4)	6.1 (1.7)	0.77	0.58	0.79
English Log frequency[b]	1.6 (0.7)	1.7 (0.5)	1.8 (0.7)	0.82	0.50	0.41
Familiarity[c] (100-700)	557.3 (41.5)	560.9 (51.1)	581.7 (41.7)	0.38	0.68	0.21
English length	5.5 (2.0)	5.5 (2.3)	5.8 (2.2)	0.95	0.62	0.54
Age of acquisition[d] (100-700)	310.6 (84.6)	304.8 (93.4)	301.7 (100.6)	0.86	0.96	0.90
Imageability[c] (100-700)	509.1 (114.4)	526.6 (88.8)	523.2 (108.2)	0.31	0.95	0.34

Note. Standard Deviations are in parentheses. [a]Quasthoff, U. (2002); [b]Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995); [c]Coltheart (1981); [d]Gilhooly, K. J. & Logie, R H. (1980)

5.1.1.3 Procedure

After completion of the translation recognition task and the operation-span task, participants were presented with bare stem nouns in all capital letters and had to decide via button press whether the given noun was masculine, feminine, or neuter. Participants used three keys on the computer keyboard (c, b, m) to indicate the gender. Key assignment mapping was kept consistent between participants. Response latencies and accuracy were recorded and analyzed.

5.1.2 Results and Discussion

Whenever the assumption of sphericity was violated, as indicated by Mauchly's test, the degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity. Incorrect responses, as well as items with response latencies below 300 ms and above 3000 ms or deviating 2.5 *SD*s from a participant's mean were excluded from reaction time analyses. Reaction times and percent accuracy scores by participant (*F*1) and item (*F*2) were entered into a one-way repeated measures analysis of variance (ANOVA) with noun type (typical gender assignment, atypical gender assignment, ambiguous gender assignment) as the independent variable.

5.1.2.1 Latencies

Results revealed a main effect of noun type ($F1(2, 162) = 48.97$, $MSE = 46125.94$, $p < .001$; $F2(2, 72) = 19.68$, $MSE = 32755.02$, $p < .001$). Pairwise comparisons using a Bonferroni adjustment indicated that participants assigned gender significantly faster to the typical nouns than the atypical nouns ($p < .001$), and assigned gender significantly faster to the atypical nouns than the ambiguous nouns ($p < .001$). Participant mean latencies are reported in Figure 5.1.

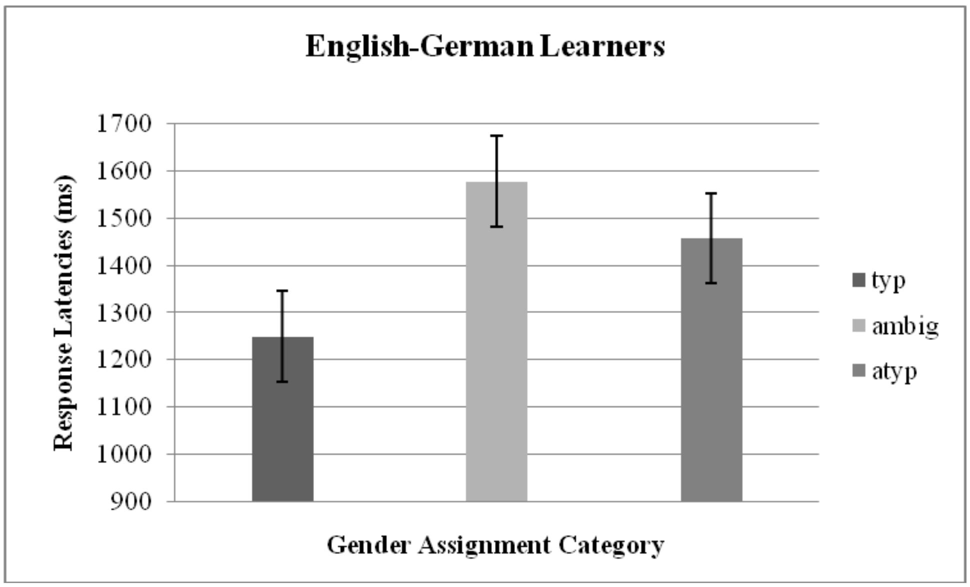


Figure 5.1: Mean gender assignment latencies for English-German learners as a function of gender category where typ = typical gender assignment, ambig = ambiguous gender assignment, and atyp = atypical gender assignment.

5.1.2.2 Accuracy

Accuracy results showed a main effect of noun type ($F1(2, 162) = 155.59$, MSE = 192.78, $p < .001$; $F2(2, 72) = 42.34$, MSE = 224.17, $p < .001$). Pairwise comparisons using a Bonferroni adjustment indicated that participants assigned gender significantly more accurately to the typical nouns than the ambiguous or atypical nouns (ps < .001).

The ambiguous and atypical nouns did not significantly differ in accuracy ($p > .1$). Participant mean latencies are reported in Figure 5.2.

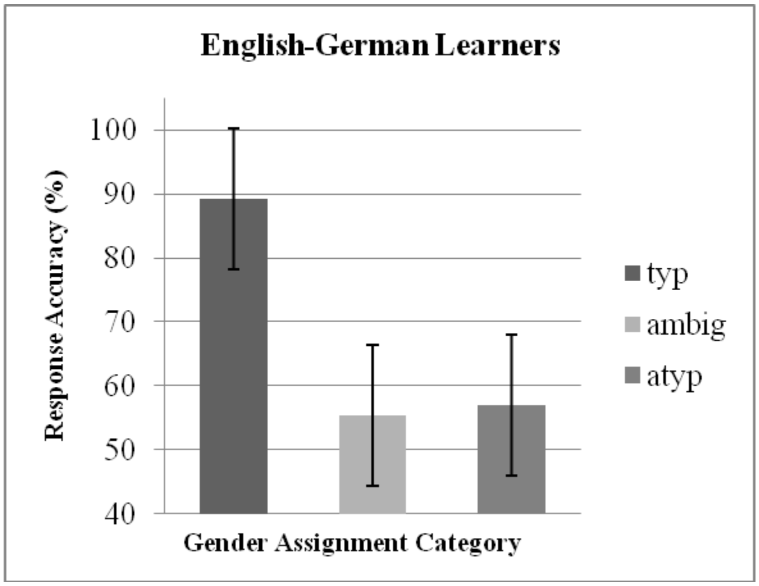


Figure 5.2: Mean gender assignment accuracy scores for English-German learners as a function of gender category where typ = typical gender assignment, ambig = ambiguous gender assignment, and atyp = atypical gender assignment.

Overall results on response latencies indicated significantly faster response latencies for typical nouns as compared to typical and ambiguous nouns. Percent accuracy scores corroborated this finding, with greater accuracy for typical nouns. Interestingly, the percent accuracy in the other two categories mirrored accuracy rates seen in the translation-recognition task. Unlike Bordag et al. (2006) in which atypical nouns were processed the slowest, in this group of learners, ambiguous nouns showed the slowest latencies. Upon closer inspection of the items included in this analysis, several of the items in the atypical category did in fact follow predictable gender rules, such as *Leitung* (faucet or pipe), *Erfahrung* (experience), and *Ahnung* (hunch) which all end in –

ung, an ending that takes the feminine gender. These items may have made distinguishing the gender of items in this category easier, thus accounting for the difference in findings with Bordag et al. Given the already small sample of words, a new analysis without these items was not performed, but in the future it would be worthwhile to conduct a new analysis replacing these three items.

In a next step, the question was asked whether this reliance on gender rules would change based on proficiency. For this purpose, participants were split into the same two proficiency groups as in Experiment 1A based on overall accuracy scores on the entire sample of gender assignment data (i.e., not just the select 75 critical items of the current analysis). While this selection criteria is not completely independent of the task, for the purposes of cross comparisons with Experiment 1, the same grouping seemed important. Reaction times on the gender assignment task as well percent accuracy scores by participant ($F1$) and item ($F2$) were entered into a 3 x 2 mixed ANOVA with noun type (typical, ambiguous, atypical) as a within-participants variable and proficiency group (high vs. low) as a between-participants variable. In the item analysis, proficiency group was a within-item variable and noun type a between-items variable.

5.1.2.3 Latencies by Proficiency

Again, there was a significant main effect of noun type ($F1(2, 160) = 51.17$, $MSE = 43619.78$, $p < .001$; $F2(2, 72) = 24.98$, $MSE = 63191.22$, $p < .001$) and a significant main effect of proficiency group ($F1(1, 80) = 30.05$, $MSE = 326069.07$, $p < .001$; $F2(1,72) = 238.97$, $MSE = 20347.26$, $p < .001$) as well as a significant interaction ($F1$(2,

160) = 5.655, *MSE* = 43619.78, $p < .01$; *F*2(2, 72) = 6.58, *MSE* = 20347.26, $p < .01$). To follow up the significant interaction, separate ANOVAs were performed for each proficiency group. Results for the low proficiency group revealed a significant effect of noun type (*F*1(2, 82) = 27.52, *MSE* = 65649.42, $p < .001$; *F*2(2, 72) = 25.04,*MSE* = 5821.30, $p<.001$). Pairwise comparisons using a Bonferroni adjustment indicated that low proficiency participants assigned gender significantly faster to the typical nouns than the atypical nouns ($p < .001$), and the ambiguous nouns ($p < .001$). There were no significant differences between the atypical and ambiguous nouns ($p > .1$). Results for the high proficiency group revealed a significant effect of noun type (*F*1(1.61, 62.85) = 34.16, *MSE* = 25392.47, $p < .001$; *F*2(2, 72) = 13.44, *MSE* = 32717.18, $p < .001$). Pairwise comparisons revealed significant differences between all three types of nouns (*p*s < .001). Participant mean latencies are reported in Figure 5.3.

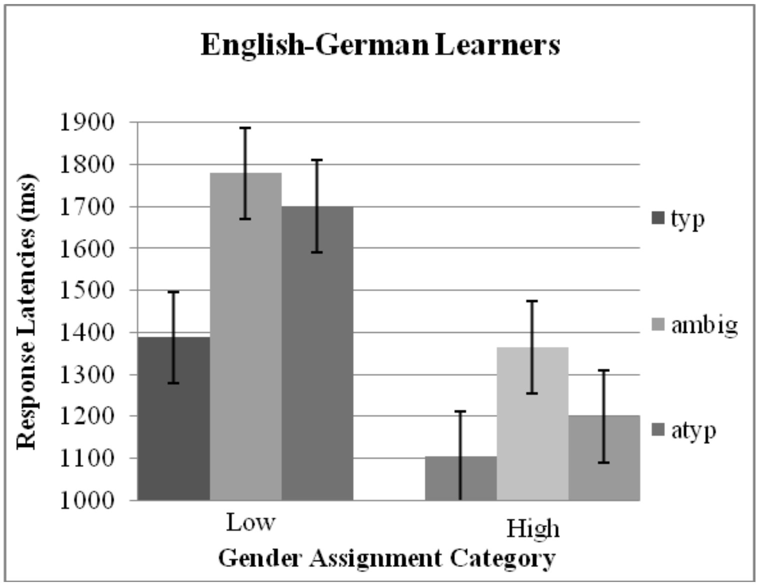


Figure 5.3: Mean gender assignment latencies for English-German learners as a function of proficiency and gender category where typ = typical gender assignment, ambig = ambiguous gender assignment, atyp = atypical gender assignment, Low = low proficiency, and High = high proficiency.

5.1.2.4 Accuracy by Proficiency

There was a significant main effect of noun type ($F1(2, 160) = 164.32$, $MSE = 180.76$, $p < .001$; $F2(2, 72) = 41.75$, $MSE = 451.03$, $p < .001$), and a significant main effect of proficiency group ($F1(1, 80) = 118.01$, $MSE = 211.18$, $p < .001$; $F2(1, 72) = 159.83$, $MSE = 91.01$, $p < .001$), but these results were not surprising given that gender assignment accuracy was the criteria by which the groups were selected in the first place. However, and more importantly, there was also a significant interaction ($F1(2, 160) = 6.39$, $MSE = 180.76$, $p < .01$; $F2(2, 72) = 6.48$, $MSE = 91.01$, $p < .01$). To follow up the significant interaction, separate ANOVAs were run for each proficiency group. Results for the low proficiency group revealed a main effect of noun type ($F1(2, 82) = 89.26$, *MSE* = 242.07, *p* < .001); *F*2(2,72) = 59.94, *MSE* = 214.58, *p* < .001). Pairwise comparisons using a Bonferroni adjustment indicated that participants assigned gender significantly more accurately to the typical nouns than the ambiguous or atypical nouns (*p*s < .001). The ambiguous and atypical nouns did not significantly differ in accuracy (*p* > .1). Results for the high proficiency group also revealed a main effect of noun type (*F*1(2, 78) = 82.06, *MSE* = 116.30, *p* < .001; *F*2(2,72) = 20.03, *MSE* = 327.47, *p*<.001). Pairwise comparisons using a Bonferroni adjustment indicated that participants assigned gender significantly more accurately to the typical nouns than the ambiguous or atypical nouns (*p*s < .001). The ambiguous and atypical nouns did not significantly differ in accuracy (*p* > .1). Participant mean latencies are reported in Figure 5.4.

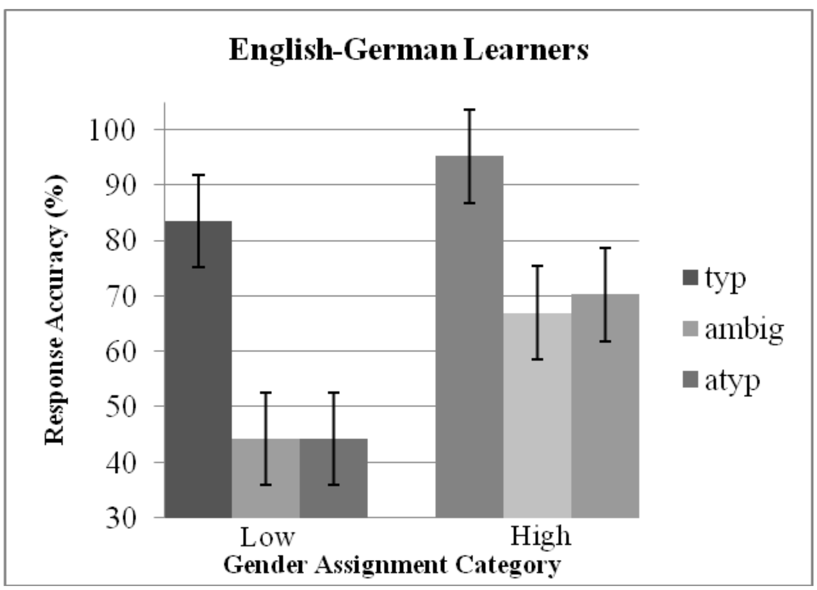


Figure 5.4: Mean gender assignment latencies for English-German learners as a function of proficiency and gender category where typ = typical gender assignment, ambig = ambiguous gender assignment, atyp = atypical gender assignment, Low = low proficiency, and High = high proficiency.

Results by proficiency revealed significant noun type by proficiency interactions in both latencies and accuracy, suggesting that, contrary to results in Experiment 1A, increased proficiency does change the way participants in this subset of L2 learners assign gender to nouns. Pairwise comparisons on accuracy for each proficiency group, however, suggests that the general pattern does not change between groups. Pairwise comparisons for latencies, on the other hand, revealed different relationships between conditions by proficiency, whereby the high proficiency group shows significant differences in latencies between all three conditions. The low proficiency group only showed differences between the typical and ambiguous conditions. In a next step, the gender assignment data from German-English learners was analyzed in order to provide a comparison for these two English-German learner proficiency groups.

5.2 Gender Assignment: German-English learners

5.2.1 Methods

5.2.1.1 Participants

The data from Experiment 1B served as the basis of the current analysis, so that data from the same 34 German-English learners in Experiment 1A was used.

5.2.1.2 Materials

Materials were the same as in the analysis of Experiment 1A.

5.2.1.3 Procedure

The procedure was the same as above.

5.2.2 Results and Discussion

Reaction times as well percent accuracy scores by participant (*F*1) and item (*F*2) were entered into a one-way repeated measures analysis of variance (ANOVA) with noun type (typical gender assignment, atypical gender assignment, ambiguous gender assignment) as the independent variable.

5.2.2.1 Latencies

Results revealed a main effect of noun type ($F1(2, 68) = 9.90$, $MSE = 3320.69$, $p < .001$; $F2(2, 72) = 3.35$, $MSE = 4802.70$, $p < .05$). Pairwise comparisons using a Bonferroni adjustment indicated that participants assigned gender significantly faster to the typical and the atypical nouns than the ambiguous nouns (*p*s < .01). There was no significant difference between the typical and atypical nouns ($p > .1$). Pairwise comparisons for items revealed significantly faster latencies for the atypical than ambiguous nouns ($p < .05$) but no other significant differences. Participant mean latencies are reported in Figure 5.5.

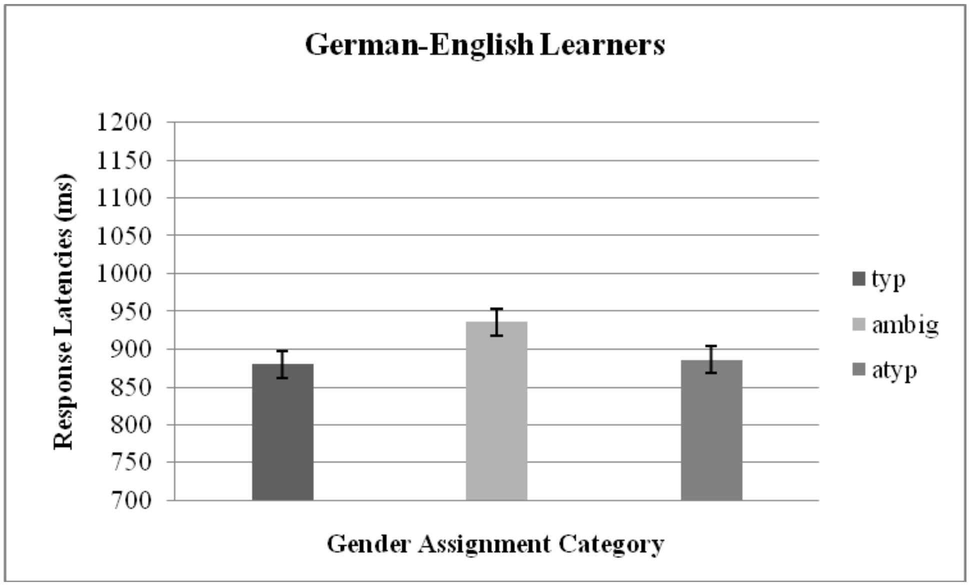


Figure 5.5: Mean gender assignment latencies for German-English learners as a function of gender category where typ = typical gender assignment, ambig = ambiguous gender assignment, and atyp = atypical gender assignment.

5.2.2.2 Accuracy

Accuracy scores were normalized using an arcsin transform. Accuracy results showed no effect of noun type ($F1(2, 68) = 1.57$, $MSE = 24.25$, $p > .1$; $F2 < 1$). Uncorrected participant means are reported in Figure 5.6.

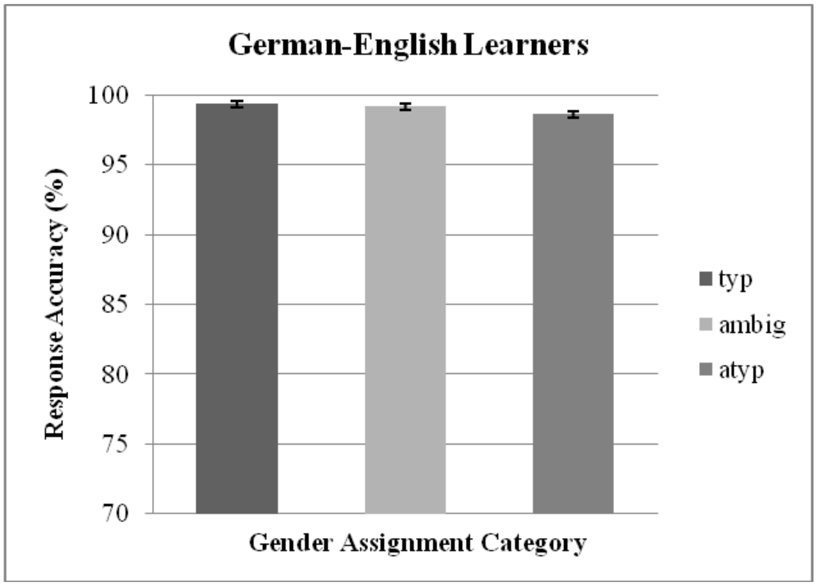


Figure 5.6: Mean gender assignment accuracy scores for German-English learners as a function of gender category where typ = typical gender assignment, ambig = ambiguous gender assignment, and atyp = atypical gender assignment.

The results suggest that German-English learners are sensitive to gender rules in their native language. Unlike the results of Bordag et al. (2006), native speakers of German did show sensitivity to phonological rules, showing significantly slower reaction times to ambiguously-marked nouns than to typical or atypical nouns. These findings, however, are in line with other monolingual research, such as Schiller et al. (2003), which did find sensitivity to phonological gender rules in native speakers. The lack of a clear distinction between typically and atypically marked nouns sheds some doubt on this interpretation but may be due, as previously noted, to the types of items included in the atypical category. Significantly, the pattern of the German-English learners approximated those of the English-German learners in their sensitivity to the typical noun condition (which had the phonological –e ending rule for feminine nouns). Unlike the German-English learners, English-German learners additionally showed a significant distinction in response latencies between the typical and atypical conditions, with longer latencies in the atypical condition. These results may suggest an increased sensitivity of English-German learners to properties such as the phonological distribution of noun endings in assigning gender in an effort to “make sense” of the seemingly arbitrary assignment of gender to nouns.

5.3 General Discussion

The overall picture that emerges from the results of both Experiment 1 and this metalinguistic task confirms previous findings that second language learners of German have a difficult time processing grammatical gender. In Experiment 1, while there were clear developmental patterns in the data in which participants responded faster and more accurately with increased proficiency, increased proficiency did not seem to change how gender was being processed: Participants continued to show particular difficulties in the condition in which they had to ignore the semantic information of the translation and focus on the accuracy of the gender. Despite improved performance, the English-German learners also did not seem to be sensitive to the gender of wrong translations when it matched the gender of the correct translation. The results of the gender assignment task shed potential light on these difficulties, revealing that nouns without a salient cue to gender pose particular challenges to L2 learners. Conversely, those nouns which do have a phonological cue for the appropriate gender are much easier to process, suggesting that learners can use features, such as phonological markings, to bootstrap their way into the gender system. While gender sensitivity is apparent in a task which is more under the learner’s control, as in the gender assignment task, more automatic processing, as in the translation recognition task, does not seem to show such sensitivity. One possibility for further investigation would be to reanalyze data from Experiment 1 as a function of whether the cues were consistent or not. This issue will be returned to in the General Discussion, in order to consider whether sensitivity to cues in a metalinguistic task might have implications for how it could be exploited in on-line processing. Another possibility is that the timing of the gender assignment task and the translation recognition task is not sensitive enough to measure developmental changes in gender processing, and this possibility will be explored in more detail in Experiment 2.

Chapter 6

Experiment 2: Simple Nouns in ERPs

Recent years have seen a rise in the use of event-related potentials (ERPs) to help disentangle the time course of language processing, although the actual technology for measuring cortical brain activity from the human scalp, electroencephalography (EEG), has existed since the work of the German psychophysiologist Hans Berger in the early 1900s. In order to understand ERPs, it is helpful to first discuss some general background to EEG. Neural brain cells communicate via electric impulses, and the collective impulse generated by a group of neurons in the cerebral cortex can then be measured at the scalp's surface with the help of electrodes placed on the scalp. Since the actual impulse that is transmitted is very small (between 5 to 10 μV), the signal must first be amplified. The recorded, amplified brain activity is called EEG (For a more detailed explanation of EEG as well as ERP, please refer to Coles & Rugg, 1997, Luck, 2005, or Molfese, Molfese, & Kelly, 2001). While the temporal precision of EEG/ERPs is very good (within milliseconds of processing), due to the physiology of the EEG, the spatial resolution (i.e., being able to locate where in the brain the activity is coming from) is very poor. This "inverse problem" as it is called is due to the fact that a given electrode picks up activity not only directly beneath it, but also from surrounding cortical areas, which complicates the localization of the activity (see Van Petten & Luka, 2006). With certain modeling techniques, however, it is possible to begin to disentangle the brain regions from which the signal originates.

It is very difficult, however, to identify the brain response to a specific item from the continuous EEG. ERPs address these constraints by taking only the part of the EEG stream that specifically relates to the processing of a stimuli/event of interest in a process called “time locking”. The signal of these data is then averaged over many trials to produce the event-related potential or ERP. The averaging technique allows random variation in neural activity which is unrelated to the appearance of the stimuli to be filtered out, leaving only electrophysical activity related to the stimuli. The peaks of the resulting weaveforms, also called components, have been related to specific cognitive processes and can be used to provide insight into the processing mechanisms of the brain.

ERP components are typically classified by their polarity (positive vs. negative going waveform), peak latency (point in time, in ms, at which the waveform peaks), and/or waveform (the order of appearance such as first negativity). A P600 would be a positivity 600 ms after stimulus onset, while an N1 would be the first negativity after stimulus onset. Components occurring early in the time course of processing, typically before 100 ms, are labeled “exogenous” and are believed to be determined by physical features of the stimulus such as frequency, color, and contrast. Components that appear later are termed “endogenous” and appear to be under the influence of the task, subject to the cognitive state of the participant. In a time window between 100 – 200 ms after stimulus onset, components may be sensitive to both physical stimulus and psychological factors (see Donchin et al., 1978, and Regan, 1989, for further discussion).

The interpretation of ERPs generally follows two alternative approaches: a physiological approach in which the anatomical source within the brain is localized, or a functional approach which focuses on the cognitive processes correlated with the ERP.

The current study will focus on a functional approach, using ERPs as a window into the cognitive processing of a second language. As such, a brief overview of potential components of particular interest and their related cognitive processes will be provided in the following section.

6.1 ERP Components of Interest

6.1.1 The LAN (250-500 ms)

The left anterior negativity (LAN) usually peaks around 250-350 ms and has a frontal or left frontal distribution. It has been found to indicate morphosyntactic violations such as gender (e.g., Deutsch & Bentin, 2001; Gunter et al., 2000; Hofmann, 2005; Koester et al., 2004) but has also been interpreted to reflect increased working memory load in a sentence context (e.g., Coulson, King, & Kutas, 1998; Kluender & Kutas, 1993). Given that the translation recognition paradigm used in the present study should make fewer demands on working memory than a sentence task, it was anticipated that LANs, if present, would reflect the detection of morphosyntactic violations. LAN effects are often registered in combination with a P600 component in a sentence context (e.g., Barber & Carreiras, 2005, Experiment 2; Hagoort & Brown, 1999; Hagoort, Wassenaar, & Brown, 2003). However, studies at the word level (i.e., not in a sentence context) have shown a LAN-N400 pattern with no P600 to gender disagreement (e.g., Barber & Carreiras, 2005, Experiment 1; Hofmann, 2005).

6.1.2 The N400 (300-600 ms)

One of the most important components implicated in lexical processing is the N400. The N400 is a negative deflection that typically peaks at 400 ms with a centroparietal distribution and seems to reflect post-lexical semantic processing (e.g., Holcomb, 1993; for a review see Kutas & Federmeier, 2000). According to Holcomb and Grainger (2006), there are several components which form a set leading up to the N400 indexing lexical processing. In specific, it is believed that the P150 is involved in the processing of elemental features in lexical processing. The effects of the P150 are relatively small, located posterior and in the right hemisphere. The N250, which has a broader distribution in more frontal sites, is believed to be involved in processing the relative position of letters. Lexical selection of single words is then indexed by the P325 which has a right hemisphere distribution. This temporal progression fits well with the combination of visual discrimination and lexical processing, moving from a feature to a more global level, which then culminates in the N400, where form and meaning interface.

At the sentence level, Hagoort & Brown (1999) investigated the syntactic nature of gender by presenting gender congruent and incongruent article-noun phrases at either the beginning or the end of a sentence. In addition to a syntactic P600 effect, results showed an N400 gender congruency effect, with an increased negativity for incongruent nouns, but only at sentence-final position, as a consequence of the syntactic violation. These results supported the idea that processing agreement is a syntactic, not conceptually, driven process. At the word level, as reviewed in Chapter 1, Barber & Carreiras (2005) found an N400 effect to gender violations in native speakers of Spanish.

6.1.3 The P300 (300-750 ms)

The P300 is a positive shift in the EEG between 300 and 750 ms and can appear either fronto-centrally or centro-parietally. It is a wide-spread component, seen in any task that requires stimulus discrimination. In so called “odd-ball paradigms” where the task involves detection of novel stimuli, the distribution is more frontal and is called a P3a (e.g., Hillyard & Picton, 1987). The P3b shows a more parietal distribution in response to stimuli that are attended and require action and also appears in response to context-updating operations (e.g., Squires, Squires, & Hillyard, 1975; for a review see Polich, 2007). The P300 may also index memory-related storage operations, showing an involvement of working memory as well as neural inhibitory mechanisms (for a recent review of these issues, see Polich, 2007). As reviewed earlier, Barber and Carreiras (2005) found delayed P3 latencies at the word level in gender disagreement compared to number disagreement and interpreted these findings as indicating a gender reanalysis process (see also Barber & Carreiras, 2003).

The picture that emerges from ERP research is that lexical processing is complex, with many different neurocognitive components at different time points with different scalp distributions. Beyond mere behavioral issues, one sees the interaction of systems, and the overlap but also specificity of various systems.

6.2 Experiment 2: Simple Nouns in ERPs

In view of the findings of Tokowicz and MacWhinney (2005), Weber-Fox and Neville (1996), Osterhout, McLaughlin, Inoue, and Loveless (2000), and most recently

Kotz, Holcomb, and Osterhout (2007), a dissociation between ERP and behavioral results was expected. Specifically, similar to the study by Tokowicz and MacWhinney investigating grammatical gender, it was anticipated that ERP results would show sensitivity to grammatical gender violations and semantic violation even though behavioral results for the English-German L2 learners had shown relatively low accuracy rates in detecting gender violations.

6.2.1 Method

6.2.1.1 Participants

10 native speakers of English learning German as a second language with intermediate proficiency volunteered to participate and were paid for their participation ($10/hour). All except one of the participants had previously participated in Experiment 1. All participants were right handed and reported being free of neurological disorders and had normal or corrected-to-normal vision. One participant's data was discarded due to technical difficulties. Table 6.1 provides language background information for these 9 participants.

Table 6.1: Characteristics of Participants in Experiment 2

	ERP Participants (n=9)
Age (years)	22.67 (2.74)
L1 self-ratings (10-pt scale)	9.6 (0.68)
L2 self-ratings (10-pt scale)	7.31 (1.06)
L2 age of acquisition (years)	13.89 (1.62)
Months of adult L2 immersion	8.22 (9.3)
Simon effect (ms)	29.72 (28.26)
Operation span (1-60)	50.38 (4.9)

6.2.1.2 Materials

162 simple nouns were distributed such that all items appeared in each of three critical conditions across three versions. The three critical conditions were the same as in Experiment 1: A. an incorrect translation item which matched the correct translation in gender, B. an incorrect translation item which did not match the correct translation in gender, and C. an item which was the correct translation but was assigned an incorrect gender. Within each version, each of the items appeared once in English as the prime of a translation pair and once in German as a target translation of another pair. Across versions, the order of language presentation (i.e., if the item first appeared as an English prime or first in German as a target translation) was counterbalanced. Within each

version, items were matched across conditions on frequency and word length in English and German, as well as English age of acquisition, familiarity, and imageability. Some of the items from Experiment 1 were used, so that accuracy on these items served as an additional variable for matching where possible. These repeated items were frequently, but not always, matched with new false translations, minimizing but not eliminating the factor of previous item exposure. Lexical characteristics of the critical items are provided in Table 6.2. For a full list of the materials, please refer to Appendix F.

Table 6.2: Item Norms

Descriptive Statistics	Mean	Std. Deviation
German frequency[a]	10.7	2.2
German log frequency[b]	1.4	0.6
German length	5.4	1.4
English frequency[c]	70.5	81.1
English log frequency[c]	1.6	0.5
English frequency2[d]	71.3	80.4
English familiarity (100-700)[e]	550.7	50.5
English length	5.2	1.6
English age of acquisition (100-700)[f]	297.9	77.7
English imageability (100-700)[e]	536.6	93.0
L2 gender accuracy (%)	58.1	22.4

Notes. [a]Quasthoff, U. (2002); [b,c]Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995); [d]Kučera, H., & Francis, W. N. (1967); [e]Coltheart (1981); [f]Gilhooly, K. J. & Logie, R H. (1980)

6.2.1.3 Procedure

The procedure was similar to the behavioral study except for a few changes in timing parameters to accommodate electrophysiological responses and minimize ocular artifacts. Stimulus presentation was as follows: Prior to each trial, a fixation sign (+) appeared at the center of the computer screen. Participants were told that they could blink when the fixation sign was visible but to avoid blinking during stimulus presentation. After they had finished blinking, they pressed a button to begin the trial, followed by an ISI of 250 ms. First an article and noun in English appeared in the center of the screen, presented word by word, followed by an article and noun in German. The English article remained on the screen for 300 ms followed by an inter-stimulus interval (ISI) of 300 ms and then the English noun for 300 ms and an ISI of 500 ms. The German article then appeared for 500 ms, followed by an ISI of 300 ms, and a German noun for 500 ms. After presentation of the German noun, participants were to indicate as quickly as possible whether the German noun phrase was the correct translation of the English noun phrase by pressing a “yes” button or a “no” button. Participants were given 3000 ms from onset of the noun to respond before the next fixation sign appeared. Response hand was counterbalanced. Participants completed 15 practice items with feedback including all three types of incorrect translations. Similar to the behavioral experiment, participants were verbally reminded to pay attention to the accuracy of the German article paired with the German noun, and to reject the translation as incorrect if the article did not agree with the noun. Additional practice trials were provided if participants had trouble with the timing or procedure.

6.2.1.4 EEG Recording and Analysis

The continuous electroencephalogram (EEG) was recorded using a 64-channel sintered Ag/AgCl electrode array mounted in an elastic cap according to the 10-20 system (QuikCap, Neuroscan Inc.). Recordings were referenced to a site between CZ and PZ during recording and re-referenced off-line to the average activity of the left and right mastoids. Electrode impedances were kept below 5 kΩ. Lateral eye movements were measured by electrodes placed on the outer canthus of each eye. Vertical eye movements were measured by electrodes placed on the upper and lower orbital ridge of the left eye. Eye recordings were later used off-line to reject contaminated trials. The electrophysiological signals were amplified using Neuroscan Synamps with a band pass filter of 0.05 to 100 Hz and a sampling rate of 500 Hz.

Only trials with correct responses were included in the analyses. A pre-stimulus baseline of 200 ms and an epoch duration of 800 ms post-stimulus were used to compute average ERPs per condition. Trials with eye movement artifacts or blinks and peak-to-peak deflections over 200 µV were rejected. A digital low-pass filter of 30 Hz (24 dB/oct) was applied when analyzing the data off-line.

Based on visual inspection, components of interest and two corresponding time windows were selected for further investigation. Mean amplitudes were calculated over each of the time windows corresponding to the component of interest, an early negativity, or N250, between 200 and 300 ms, and a late negativity, or N400, between 300 and 600 ms. Two sets of analyses were conducted for each time window, one on three representative electrodes for the midline sites (FZ, CZ, and PZ) and one on representative electrodes for the lateral sites (left hemisphere: F3, C3, and P3; right hemisphere: F4, C4, and P4) in which each electrode corresponded to a frontal, central, or parietal site. The relative location of these electrodes can be seen in Appendix G. For each dependent variable (i.e., mean amplitudes of a given component), a 4 (translation condition: right article, right noun; right article, wrong noun; wrong article, right noun; wrong article, wrong noun) X 3 (electrode: frontal, central, parietal) ANOVA was conducted. For the lateral sites, the additional variable of hemisphere (left vs. right) was included. For the variable of translation condition, the "right article, right noun" condition was included even though this condition constituted "yes" responses, while the other three conditions constituted "no" responses. Inclusion of this fourth variable allowed exploration of the perceived grammaticality of the "wrong article, right noun" condition which may pattern more like "yes" responses if participants are not sensitive to grammatical gender violations for correct noun translations. If anything, inclusion of this fourth variable constituted a conservative approach, even though it represented a different type of response. Typically, a larger P3 would be expected for "no" responses, which might be expected to "pull down" or attenuate the N400. If this attenuation does in fact occur, then it would make it more difficult to find any effects on the N400. Violations of sphericity of the data were corrected using the Greenhouse-Geisser correction (Greenhouse & Geisser, 1959). Follow-up comparisons for significant effects will only be performed for condition or interactions of electrode or hemisphere with condition, since the main interest in the present experiment is how translation conditions differ as opposed to how ERPs pattern more generally. While data were collected from both the onset of the German article and the German noun, the following analyses will focus on the noun presentation. Only the EEG results, and not behavioral results, will be reported here as these are of primary interest to the current experiment, and these data are of an exploratory nature. ERPs elicited by critical nouns at representative electrode-sites are presented in Figure 6.1. For ERPs from the full 64 electrodes, please refer to Appendix G.

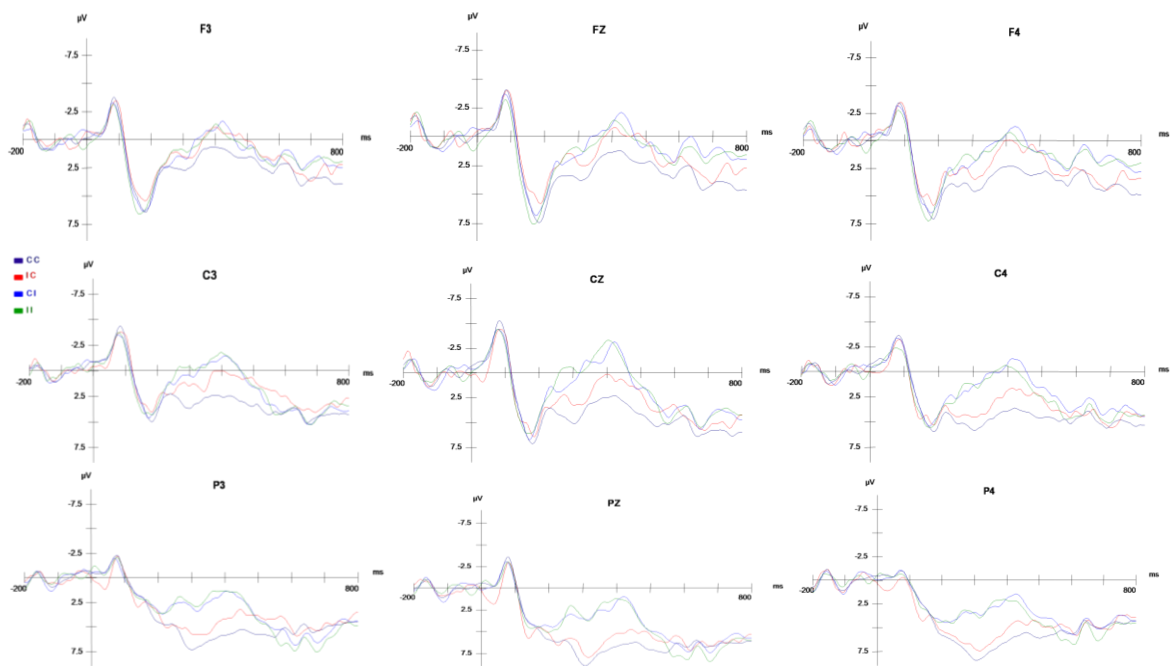


Figure 6.1: ERP waveforms at all 9 electrode sites for all four conditions, where CC = right article, right noun; IC = wrong article, right noun; CI = right article, wrong noun; and II = wrong article, wrong noun. Note negative plotted up.

6.2.2 EEG Results

Visual inspection of the components revealed a general pattern of an initial negative peak at 100 ms, consistent with an N100, followed by a positivity or P2 maximally peaking just before 200 ms. These components were followed by two additional negative peaks. The first of these negative peaks occurred approximately 250 to 300 ms into stimulus processing, consistent with an anterior negativity or possibly an N250. While the distribution of effects seems to concentrate over the left hemisphere and central sites, similar patterns are also present over the right hemisphere. The N250 is also followed by a negativity at around 400 ms, consistent with N250 – N400 pairings documented by Holcomb and Grainger (2006, see discussion above). While the earlier component could be characterized as a left anterior negativity (LAN), because of its broad distribution, and also the following negativity around 400 ms, the earlier negativity is hereafter referred to as an N250. According to the features of the second negativity, it is termed as N400, which was followed by a positivity around 600 ms with a centro-parietal distribution, consistent with a response P3. This positivity, however, was not very distinct and seems to be influenced by the preceding N400.

6.2.2.1 N250

In the analysis by hemisphere, there was a significant main effect of condition (F (1.90, 15.20) = 6.17, MSE = 13.29, p < 0.05) and hemisphere (F (1, 8) = 9.00, MSE = 8.83, p < 0.05), and a significant interaction of condition x electrode (F (2.44, 19.54) = 3.98, MSE = 3.73, p < 0.05). To further investigate the effects of condition, three one-way ANOVAs were run on the average amplitudes over hemispheres for electrodes at frontal, central, and parietal sites. Results revealed significant differences at central sites (F (1.89, 15.13) = 5.49, MSE = 2.80, p < 0.05) and parietal sites (F (1.95, 15.60) = 12.41, MSE = 2.39, p < 0.01) but not frontal sites (F (1.98, 15.81) = 1.25, MSE = 3.65, p >.1).

Pairwise comparisons using a Bonferroni correction revealed larger amplitudes at central sites in the right article, right noun condition than the right article, wrong noun condition ($p < .05$) and the wrong article, wrong noun condition ($p < .05$). No other differences were significant (*p*s > .1). Similar to the comparisons at central sites, pairwise comparisons using a Bonferroni correction revealed larger amplitudes at parietal sites in the right article, right noun condition than the right article, wrong noun condition ($p < .01$) and the wrong article, wrong noun condition ($p < .01$). Additionally, there was also a significant difference between the wrong article, right noun condition and the wrong article, wrong noun condition ($p < .05$). These comparisons suggest the influence of semantic processing on translation, but do not indicate a strong influence of gender sensitivity in translation at lateral sites, although possibly a hint of an effect of gender mismatch.

In the analyses for midline electrodes, results revealed a main effect of condition (F (2.01, 16.09) = 7.14, MSE = 8.22, $p < 0.01$) but no main effect of electrode (F (1.12, 8.93) = 2.90, MSE = 38.29, $p > 0.1$). There was, however, a significant condition x electrode interaction (F (2.29, 18.31) = 5.13, MSE = 2.18, $p < 0.05$). To further investigate the interaction, separate ANOVAs were performed for each electrode site (FZ, CZ, PZ) revealing significant effects of condition at CZ (F (2.10, 16.77) = 8.23, MSE = 3.10, $p < 0.01$) and PZ (F (1.67, 13.30) = 11.26, MSE = 4.02, $p < 0.001$) but not FZ (F (2.07, 16.56) = 1.80, MSE = 4.03, $p > 0.1$). Pairwise comparisons using a Bonferroni correction revealed more positive amplitudes at CZ in the right article, right noun condition than the right article, wrong noun condition ($p < .05$) and the wrong article, wrong noun condition ($p < .01$). At PZ, identical to CZ, comparisons revealed more positive amplitudes in the right article, right noun condition than the right article, wrong noun condition ($p < .01$) and the wrong article, wrong noun condition ($p < .001$). These results, similar to the results by hemisphere, suggest that participants are sensitive to the semantic incongruence of the translations but not the gender mismatch.

6.2.2.2 N400

In the analysis by hemisphere, there were significant main effects of condition (F (1.83, 14.60) = 6.95, MSE = 17.27, $p < .01$) and electrode (F (1.04, 8.36) = 7.13, MSE = 58.35, $p < .05$), but not of hemisphere (F (1, 8) = 3.06, MSE = 12.66, $p > .1$). The main effect of condition was qualified by a significant condition x hemisphere interaction (F (2.94, 23.48) = 6.12, MSE = .66, $p < .01$). To follow up the significant condition x hemisphere interaction, separate ANOVAs were conducted for amplitudes averaged over electrodes in each hemisphere. Results revealed significant effects of condition for both left and right hemispheres (left: F (1.86, 14.89) = 5.29, MSE = 2.68, $p < .05$; right: F (1.89, 15.09) = 8.21, MSE = 3.27, $p < .01$). Pairwise comparisons using a Bonferroni correction for the left hemisphere revealed significantly more positive amplitudes in the right article, right noun condition than the right article, wrong noun condition ($p < .05$). Amplitude differences between the right article, right noun condition and the wrong article, wrong noun condition approached significance ($p < .1$). Pairwise comparisons for the right hemisphere revealed significantly more positive amplitudes in the right article, right noun condition than the right article, wrong noun condition ($p < .05$) and the wrong article, wrong noun condition ($p < .05$). Like the previous results, these results seem to support semantic awareness in the second language learners, but not sensitivity to grammatical gender.

In the analyses for midline electrodes, results revealed a main effect of condition (F (2.03, 16.27) = 8.84, MSE = 9.22, $p < 0.01$) as well as a main effect of electrode (F (1.13, 9.02) = 7.63, MSE = 42.63, $p < 0.05$), but no interaction (F (2.32, 18.55) = 2.61, MSE = 2.37, $p < 0.1$). Pairwise comparisons for the main effect of condition using a Bonferroni correction revealed significantly more positive amplitudes in the right article, right noun condition than the right article, wrong noun condition ($p < .05$) and the wrong article, wrong noun ($p < .05$). Amplitudes in the wrong article, right noun condition and the wrong article, wrong noun condition also differed significantly ($p < .05$) with more positive amplitudes in the former condition. Pairwise comparisons for the main effect of electrode using a Bonferroni correction revealed marginally more positive amplitudes for PZ over FZ ($p < .1$) and CZ ($p > .1$). These results, particularly the comparison of the conditions, again underscore the role of semantics in the processing of translations, and the absence of an obvious gender effect, at least in this sample of participants.

In summary, significant differences between conditions were found both in the 200-300 ms (N250) time window and the 300-600 ms (N400) time window. The general trend was for significant differences between the “yes” trials (right article, right noun translation) and the two “no” conditions in which the noun was incorrect (the right article, wrong noun condition and the wrong article, wrong noun condition). However, there was no significant difference between these latter two conditions, suggesting that the significant difference between the “yes” trials and these two “no” trial conditions was the difference in semantics. The lack of a significant difference between these two conditions is of particular importance, as a difference in sensitivity to these two conditions could have indicated a sensitivity to gender. There was also no significant difference between the right article, right noun condition and the wrong article, right noun condition, indicating that these two conditions were processed similarly. In essence, even though the former was a “no” condition (to which participants responded “no” correctly) and the latter a “yes” condition, participants seemed to have perceived the wrong article, right noun condition as a correct translation.

6.3 General Discussion

Statistically there seems to be little support for gender sensitivity in this sample of second language learners of German; however, visual inspection of the components at midline sites suggests that there may be hints of an effect of gender processing. Please refer to Figure 6.2

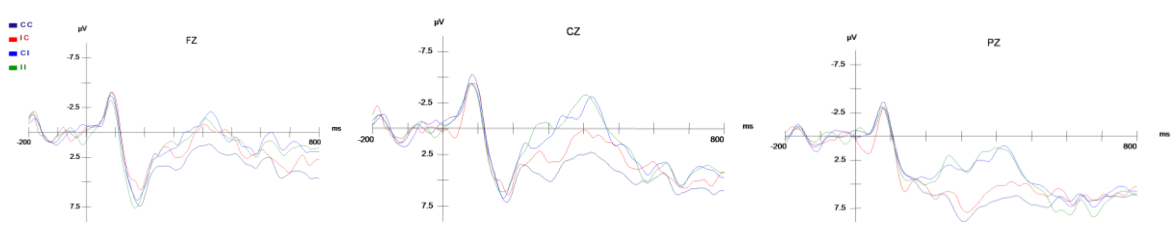


Figure 6.2: Waveforms for all four conditions at electrode sites FZ, CZ, PZ. Note negative plotted up.

While the right article, wrong noun condition and the wrong article, wrong noun condition seem to pattern very similarly, there is a reduction in the amplitude of both the N250 and the N400 for the wrong article, right noun condition. There is an even further reduction in amplitude for the right article, right noun condition. If the effect had been purely semantic, as suggested by the statistical analysis, one would have expected the components in these two conditions to show few differences, similar to the other two conditions. The modulation of the components suggests that the wrong article, right noun condition is not actually being processed identically to the correct translations. To further explore this possibility, correct items from the wrong article, right noun condition could be plotted against incorrect items from the same condition. Data from the “yes” trials could also be plotted for comparison.

Due to the exploratory nature of the study at this point, it is not possible to conclusively state whether with increasing proficiency these waveforms might begin to significantly diverge between conditions. Clearly evidence from a group of native speakers of German who are learners of English is necessary as a comparison to the German learner group in order to see the effect of gender processing in each of these conditions.

Chapter 7

Experiment 3: Gender Assignment in Compounding

Results from the previous chapters suggest that there are constraints to late L2 acquisition of morphosyntax. The results from the gender assignment task, however, allow us to entertain the possibility that there may be sensitivity to gender processing in late L2 learners when there is consistency in the underlying structure. This chapter focuses on a different aspect of morphology, compounding in German, which on the one hand is more complex in that it has two (or more) components, but on the other hand is also more regular with more systematicity: Assigning gender to German compounds made up of two or more nouns follows the rule that the gender of the entire compound takes its gender from the final (head) constituent. If the constraints seen in the previous experiments are due to irregularity (i.e., the arbitrariness of the gender assignment), then late L2 learners of German may not generalize this inability to compounding. If, however, the constraints are about morphosyntax in general, and given that participants will be processing two items in parallel, one may see similar or new manifestations of constraints in compounding. The following experiment was conducted together with Experiment 1, providing the advantage of an entirely within subject design which will enable the examination of the relationship between the processing of grammatical gender and compounding. Given the rich and lively debate in psycholinguistic literature about the representation of multimorphemic words and its relevance to bilingualism, before presenting the actual experiment, I will embark on a slight digression in order to provide additional theoretical background to compounding.

7.1 Background to Compounding

It has been argued that compounds provide a window into the first word formation process as a type of "protolinguistic fossil" (Jackendoff, 2002, p. 250). Unlike abstract morphemes that are less transparent in meaning, the meaning of the individual constituents of a compound and the relationship between the constituents tend to be more apparent. A "blackbird" is therefore a bird which is black in color. In contrast, "cats" represents more than one cat, but "s" does not by itself transparently reflect plurality. Linguists hypothesize that in the development of language, words were initially more concrete, and multimorphemic words tended to be transparent compounds before language, and morphemes, became more opaque and the relationship between the sign and the signified arbitrary. The compound has therefore become a representative of the initial multimorphemic process and a tool for investigating lexical processing mechanisms in general (Libben, 2006). Recent evidence from masked-priming studies in the monolingual domain further supports the idea that compounds behave like morphologically complex words (e.g., Fiorentino, 2006; Shoolman & Andrews, 2003).

Evidence from speech errors in compounding further suggests that compounds engage general lexical processing mechanisms (Stemberger, 2001). While speech errors in compounds are rare, when they do occur, they pattern similar to other lexical speech errors. For instance, constituents of a compound, just like two adjacent nouns, may exchange positions. Thus a speaker might produce ‘lidboxes’ instead of ‘boxlids’. Similar to when adjacent words are switched, constituents of a compound exhibit stranded stress, such that the error compound and the target compound have the same stress pattern. In the example above, the primary stress would thus be on ‘lid’ in the error compound and on ‘box’ in the target compound. In this manner, compound errors seem to display a “syntax-like” processing of compounds (Stemberger, 2001, p. 439).

A general requirement of compounds is that constituents be free morphs. This means that each constituent of a compound can stand alone as a word. (An exception to this would be so called “cranberry morphs” (Aronoff, 1976) in which one of the constituents does not have transparent meaning). As a result, issues which arise for general research in word processing must also be taken into consideration for compound processing. At the same time, constituents are morphemes, so that morphological constraints will also play an important role in processing. In the sections that follow, some of the research relevant to the compounding experiment will be highlighted.

7.2 Morphological Access

A crucial question is how compounds, and morphologically complex words in general, are represented in the mind. In general there are two competing claims about how morphology is accessed. One maintains that compounds are represented as single words, and in another, compounds are decomposed into their individual constituents or morphemes. Non-compositional “whole word” theories such as Butterworth’s (1983) Full Listing Model, and connectionist theories, such as Seidenberg and McClelland’s (1989), describe morphological access in which initial activation occurs at the whole word level and morphological constituents play no role in word processing (see also Bybee, 1995; McClelland et al. 1986). On the other extreme are compositional theories (Dell, 1986; Levelt et al., 1999; Libben, et al., 1999; Taft, 2004; Taft & Forster, 1976) in which there is initial activation at the whole word level with subsequent activation of transparent morphemes (late decomposition) or initial activation of morphemes followed by whole word activation (early decomposition). Within these models, several factors can modulate activation such as semantic transparency in which only semantically transparent words are decomposed (Marslen-Wilson et al., 1994; Schreuder & Baayen, 1995) and frequency where only infrequent words, and new forms that are not represented in the mental lexicon, are decomposed (Caramazza et al, 1988).

Dual Route Models (Isel, Gunter, & Friederici, 2003; Marcus et al., 1995; Pinker & Prince, 1988) propose an intermediate position which takes both a lexical (full form) route and a constituent (morphological parsing) route into consideration. If they are activated in parallel, a race between the two routes ensues in which factors such as frequency, morphological productivity, and semantic transparency determine the outcome of the race. Alternatively, each of the routes might be activated by a different word type, in which familiar words are processed by a lexical route and unfamiliar words by a constituent route (see also Baayen et al., 1997; Isel et al., 2003; Zwitserlood, 1994).

Partial support for a Dual Route Model comes from studies such as Koester et al. (2004), in which participants were acoustically presented with two-constituent compounds. In the first experiment, participants were asked to judge the gender agreement between a determiner and the compound or compare the compound to a visually presented word on a semantic basis. Even though gender in German agrees with the right (head) constituent, gender incongruities with both the left and the right constituent elicited left-anterior negativities (LANs), supporting morphosyntactic decomposition. In contrast, in a second experiment, when linking elements were tested, number violations only elicited an N400 effect for head constituents, a finding which disagrees with predictions made by full-parsing models.

Research in the bilingual domain also seems to support a dual-route mechanism for morphological processing, identical to the one used by monolinguals. Hahne, Mueller, and Clahsen (2006) investigated regular and irregular inflected participial inflection and noun plurals in bilinguals, expecting different brain activations for violations of irregular and regular forms. By this logic, irregular forms are stored as full-forms in the lexicon so that misapplication of the irregular inflection would produce a lexical violation similar to a pseudoword. In contrast, misapplication of a regular inflection would be seen as combinatorial violations. ERP results confirmed findings in the monolingual domain, where violations of regular patterns revealed a morphosyntactic P600 and violations of irregular inflections a lexical N400.

More recent evidence further suggests that bilinguals might rely more on a full-storage route than native speakers. Silva and Clahsen (2008) investigated regular past tense forms as well as deadjectival nominalizations (e.g., words ending in –ness or –ity) in a series of masked priming experiments. Results showed that unlike in native speakers, inflected and derived word forms produced no, or reduced, priming in adult L2 learners, supporting the idea that L2 learners store morphologically complex words as whole lexical entries rather than decomposed units. These findings were particularly striking for the regulary inflected words, as these tend to show robust priming effects in L1 and are thought to benefit most from a decomposed representation.

7.3 Frequency Issues

Frequency effects are closely related to the discussion of morphological access. Indeed, several researchers have suggested that word frequency modulates whether there is full or partial word form processing (e.g., Lehtonen & Laine, 2003). In a previous study using Finnish, which is a highly inflected language, Lehtonin and Laine (2003) found that balanced Finnish-Swedish bilinguals decompose low, medium, as well as high frequency words while monolinguals treated high frequency words as full forms. They postulated that by virtue of being bilingual, participants were not exposed enough even to high frequency words to develop full form entries. These results contrast strongly with the previously cited research by Silva and Clahsen (2008) who argue for full-form listing preferences in L2 learners. Lehtonen et al. (2006) followed up on this study by comparing Swedish monomorphemic and multimorphemic words in a study investigating morphological effects in a visual lexical decision task. Swedish, unlike Finnish, is a more morphologically limited language, thus allowing them to disentangle whether previous results were due to bilingualism (i.e., a decrease in exposure to each of the languages) and whether morphological richness of a language would modulate these effects. In contrast to their previous study, Lehtonen et al. found that bilinguals decomposed only low-frequency words; in this morphologically less complex language, bilinguals did indeed develop full form representations of words. They concluded that in addition to frequency, morphological complexity of a language plays into the type of morphological access in bilingualism (see also Portin, Lehtonen, & Laine, 2006 for similar results).[6]

Frequency effects of constituents have, however, been taken into consideration in compound research. Using reaction time and eyetracking in naming, lexical decision, and sentence reading tasks, Juhasz et al. (2003) manipulated the relative frequency of each of the constituents in English compounds while keeping the overall frequency of the compounds constant. They found that it was the frequency of the final constituent which determined the effectiveness of the processing of the compound. Juhasz and colleagues proposed that the ending lexeme plays a crucial role in accessing the meaning of a compound, and, if not accessed before, is coactivated with the full compound meaning. Eye-tracking research by Andrews et al. (2004) further suggests that the elements of a compound are activated in retrieving the whole compound word.

An additional factor which must be considered is concreteness (Crutch, 2006; Crutch & Warrington, 2005). Crutch and Warrington reported a fundamental difference in how abstract and concrete words are organized in the mental lexicon in a recent case study of a patient with semantic refractoriness access disorder. Patients with this disorder appear to be unable to understand simple verbal instructions but upon introducing a delay of 15 to 20 seconds, vastly improve their performance. Whereas concrete words are

---

[6] Regarding frequency effects in morphology, however, one must be aware of two types of frequency which can influence morphological patterns. Important for the proposed experiments which examine German is that in German it appears that phonological frequency takes precedence over morphological frequency (Clahsen et al., 1993; Marcus et al.,1995; Stemberger, 2001). Thus, even though perfect –*t* and –*en* have the same frequency as suffixes, -*t* is argued to be the regular suffix because it generally does not add a syllable to a word. In German, shorter words are more frequent than longer words, indicating a preference to keep words short. There appears to be little research that considers the consequence of frequency type (phonological vs. morphological) for the processing of compound nouns.

organized via semantic category, abstract concepts are organized by semantic association. In the patient of this case study, abstract concepts appeared to be immune to relatedness effects that are typically found with concrete words but did show semantic association effects. In contrast, concrete words showed the expected semantic similarity effect, but were not affected by semantic association. Crutch and Warrington interpreted this double dissociation as a qualitative difference in the organization of abstract and concrete knowledge within the mind. Tokowicz and Kroll (2007) further showed in a sample of second language learners that the ambiguity of translations and of word meaning is correlated with concreteness, which may suggest processing implications that are likely to extend to compounds as well.

Closely related to the idea of abstract and concrete knowledge is the role of transparency in word formation. There is some evidence, however, that in production studies, while morphology is important in speech planning, semantic transparency does not seem to play a role in morphological preparation (Roelofs & Baayen, 2002). In a picture-word interference (PWI) study specifically investigating compound nouns, Dohmes, Zwitserlood, and Bölte (2004) found that morphologically related distractors facilitated picture naming, regardless of transparency, providing corroborating evidence that semantically transparent and opaque words share morphemic representations.

Gumnior et al. (2006) note that previous production studies have been largely limited to paradigms such as picture-word interference (PWI), which, by the nature of the task, require the use of concrete objects. This approach to production poses obvious limitations to the results found in these studies and, similar to the argument advanced by Crutch and colleagues (2005) discussed above, points to a subset of the lexicon which has not been tested in speech production, namely abstract words. For these reasons, Gumnior et al. (2006) introduced the use of a translation Stroop task with distractor words (cf. La Heij et al., 1990). Results suggested that findings using a word translation task with an increased subset of abstract words had comparable results to traditional PWI studies, but allowed the use of a larger and more varied set of materials. We followed a similar research strategy in the current experiment by using a translation recognition task, as discussed in Chapter 2.

A final consideration for bilinguals is the cognate status of compound constituents. It is possible that one or both constituents of a compound are cognates across the languages under investigation in the proposed studies (English and German), and this situation poses processing costs and/or benefits which may be unique to cross-linguistic situations. The use of cross-linguistic cognates to investigate general morphological processes has become increasingly prevalent. Indeed, empirical evidence supports the idea that cognate words pattern very similarly to within-language morphologically related words (Sánchez-Casas & García-Albea, 2005). Particularly in the domain of primed word recognition, research suggests that cognates seem to be represented similarly to morphologically related words at the lexical level (Sánchez-Casas & Almagro, 1999). However, this research also suggests that cognates might pose a special type of morphological relationship. It is important to note that these studies have largely used language pairings in which the languages have a shared script, such as Catalan and Spanish (e.g., García-Albea et al., 1998; Sánchez-Casas et al., 2000). Thus it is difficult to determine whether differences between cognates and other morphological effects are due to the lexical status of cognates per se, or the additional factor of an overlap in orthography/form (although see Gollan et al., 1997, for task-related issues that arise with script differences across target and prime). It is clear, however, that form similarity in and of itself does not define the special status of cognates (e.g., García-Albea et al., 1996).

As reviewed above, there is increasing evidence for morphological decomposition in monolinguals and limited evidence for decomposition in bilinguals as well. The following experiment is an initial attempt to address the representation and processing of compounds and use of grammatical gender in late L2 learners of German. As an initial investigation, and given the range of proficiency tested and the resulting limitation in available materials (i.e., lower proficiency participants know fewer compounds), variables such as world length, frequency, and AoA were not orthogonally manipulated, and instead were kept constant across conditions. By keeping these factors constant across conditions, and by using relatively familiar compounds, there was a bias, if anything, towards whole word processing. In Koester et al. (2004), the finding in Experiment 1 that native speakers of German were sensitive to the gender of the initial (non-head) constituent of a compound is somewhat surprising, particularly since other studies have suggested that the very prosody of a word indicates that the upcoming word is a compound (Isel et al., 2003) and because the gender of the initial constituent is irrelevant in determining the gender of the compound. Prosody could act as a cue to gender agreement, encouraging whole word processing, although the results of Koester et al. suggest that apparently it doesn’t, at least not in the initial processing of the compound. Previous research from the visual domain in monolinguals has also suggested that compounds are decomposed (e.g., Zwitserlood, 1994). However, little is known about how L2 learners of a gendered language represent and subsequently process compound words. Analogous to auditory processing, visual processing, as in the following experiment, contains a cue to compound status in the relatively longer length of the word in comparison to simple nouns. This cue could similarly act as a cue to encourage whole word processing in L2 learners. From a frequency point of view, L2 words of are of general lower relative frequency for L2 learners, which should encourage decomposition along the arguments that frequently used words are stored holistically for ease of processing, while less frequently used words can be computed with relatively little cost to the processing system.

One possibility is that proficiency might modulate processing, so that with increasing proficiency, the functional frequency in the L2 increases, leading to higher functional frequency of the compounds and an increased likelihood of whole word processing. Participants with lower L2 proficiency on the other hand should be more likely to decompose compounds. One problem in teasing apart these predictions that needs to be kept in mind is that lower proficiency participants are also less likely to know the gender of words, so that if they do not show sensitivity to constituent gender, one cannot be sure whether this is due to lack of gender knowledge or processing differences.

7.4 Experiment 3A

7.4.1 Method

7.4.1.1 Participants

The same 82 participants as in Experiment 1A participated in the current experiment as trials for this experiment were intermixed with Experiment 1A. This approach provided an entirely within-participant context for both experiments within which to examine the relationship between gender processing and compounding.

7.4.1.2 Materials

60 critical compounds were selected and distributed as follows: German has a gender system in which nouns are marked for masculine (der), feminine (die), or neuter (das) gender. Compound nouns take their gender from the final noun so that while "wine" is masculine ("der Wein") and "glass" is neuter (das Glas), the compound noun "Weinglas" is neuter (das Weinglas). By virtue of gender distribution across constituents within a compound, two types of compounds exist – agreement-agreement (AA), where the gender agrees with both the first and second constituent, and violation-agreement (VA), where the gender disagrees with the first constituent, but agrees with the second constituent. Critical compounds were selected by choosing noun-noun compounds, half of which had constituents that matched each other in grammatical gender (e.g., der$_{masc}$ Regen$_{masc}$ wald$_{masc}$), and half that did not match each other in gender (e.g., der Spiel$_{neut}$

platz$_{masc}$). As a result, a total of 24 cognate and 36 noncognate compound nouns were evenly divided into matched AA and VA lists, for a total of 60 critical compounds. In addition, critical items had a close one-to-one matching across language translations so that the translation of the first constituent in German mapped onto the first constituent in English, and the translation of the second constituent in German onto the second constituent in English as closely as possible (e.g., der Vogelkäfig, the bird cage). Items which did not have this one-to-one translation relationship were included in the fillers.

A particular consideration in choosing critical items was the issue of zero plurals and other case forms of gender. A given form of an article such as *die* can match onto not only *feminine* but also *plural*. In most cases, the plural form of the noun changes in a distinct way (e.g., die Katze$_{cat}$, die Katzen$_{cats}$), but in some cases, no change to the noun takes place (e.g., das Zimmer$_{room}$, die Zimmer$_{rooms}$). Similarly, *der* can either indicate nominative masculine singular, or genitive feminine singular/genitive plural (die Bürste$_{nom}$, der Bürste$_{gen}$). Care was taken to choose items so that in the violation conditions (AV, VV, VA), the presented article could not be a plausible gender for a zero plural or a gender in another case form.

Lists of AA and VA were matched as a group based on Leipzig Frequency (Quasthoff, 2002) followed by word length for the total compound, constituent frequency and word lengths and frequency in English (Kučera & Francis, 1967). The items were generally matched item-by-item as closely as possible and then across participant lists and conditions. Table 7.1 shows mean values for each compound group on each of these properties. For the full set of items, please refer to Appendix H. Some compounds insert linking elements between constituents. These were avoided if adding the linking element created a plural form of the constituent. Since some items carry not only grammatical but also biological gender, items with biological gender were excluded, including jobs such as ‘die Geschäftsfrau’. Only singular nouns were used.

Table 7.1: Characteristics of Critical Compound Items in German and English

		CompoundType AA	CompoundType VA	Significance
German Compound	Class Frequency Leipzig	14.70	14.83	0.80
		(2.25)	(1.78)	
	CELEX Log Frequency	0.22	0.29	0.61
		(0.38)	(0.47)	
	Word Length	10.97	10.47	0.40
		(2.24)	(2.34)	
Constituent 1	Leipzig Frequency	10.07	10.50	0.42
		(1.89)	(2.24)	
	CELEX Log Frequency	1.50	1.45	0.71
		(0.65)	(0.55)	
	Word Length	5.20	5.20	1.00
		(1.16)	(1.69)	
Constituent 2	Leipzig Frequency	10.77	10.43	0.65
		(2.47)	(3.15)	
	CELEX Log Frequency	1.29	1.42	0.48
		(0.69)	(0.67)	
	Word Length	5.77	5.17	0.18
		(2.03)	(1.29)	
English Compound	Word Length	9.90	10.27	0.61
		(2.22)	(3.27)	
Constituent 1	Word Length	5.13	5.17	0.94
		(1.55)	(2.00)	
	KF Frequency	99.50	108.27	0.83
		(148.26)	(173.94)	
	CELEX Frequency	119.25	129.96	0.82
		(177.46)	(181.01)	
	CELEX Log Frequency	1.70	1.82	0.65
		(0.62)	(0.42)	
	Familiarity	578.56	582.19	0.76
		(48.06)	(37.67)	
	Age of Acquisition	275.84	277.26	0.96
		(91.31)	(87.10)	
	Imageability	572.15	576.56	0.82
		(75.96)	(66.29)	
Constituent 2	Word Length	4.73	5.10	0.41
		(1.53)	(1.86)	
	KF Frequency	105.82	139.83	0.44
		(138.74)	(187.31)	
	CELEX Frequency	166.96	144.68	0.78
		(384.66)	(185.97)	
	CELEX Log Frequency	1.78	1.85	0.64
		(0.60)	(0.60)	
	Familiarity	558.83	572.81	0.29
		(54.42)	(43.34)	
	Age of Acquisition	264.55	283.05	0.50
		(75.71)	(100.89)	
	Imageability	532.31	544.93	0.59
		(85.97)	(87.65)	

Two lists were created where the critical compounds were assigned a different determiner in the second list. Items which were previously in the AA condition (der Regenwald) now were part of the VV condition (*das Regenwald). Similarly, items in the VA condition (der Spielplatz) were assigned a determiner which agreed with the first constituent, creating the AV condition (*das Spielplatz). An item which appeared in the AA condition and VA condition in the first list would appear in the VV condition and AV condition in the second list, and vice versa. For a given list, half of the critical items would appear in the "yes" condition, and half would appear in the "no" condition. Fillers were therefore created, 24 cognate and 36 noncognate compounds, where half would require a "yes" response and half would require a "no" response, this time based not on gender agreement but based on translation accuracy (e.g., THE MAILBOX – DER ESSTISCH$_{\text{dining table}}$). The same compound fillers were used for both lists and contained equal numbers of AA and VA items. Compound fillers were matched on length and frequency to critical items across lists and conditions, but not individually matched item-by-item. In addition, 112 single nouns were added as fillers for Experiment 1. None of the individual constituents were repeated as constituents of another compound either in the critical items or in the filler items, and none of the constituents appeared as simple nouns in the filler items. In this way, participants only saw a given constituent once in the entire experiment.

7.4.1.3 Procedure

The same procedure as in Experiment 1A was used. Of particular importance is the fact that compound nouns were presented visually as a whole on the screen (e.g., ESSTISCH, not ESS then TISCH).

7.4.2 Results and Discussion

7.4.2.1 Yes Trials

Analyses of the "yes" trials showed a main effect of compound type in the participant analysis ($F1(1, 81) = 10.29$, $MSE = 10351.57$, $p < .01$; $F2(1, 58) = 2.95$, $MSE = 23833.86$, $p < .1$), with significantly faster participant latencies in the violation, agreement condition, VA, than in the double agreement condition, AA ($M1$: VA = 1221 ms, AA = 1272 ms). Analyses for accuracy data confirmed these results, with a main effect of compound type in the participant analysis ($F1(1, 81) = 9.09$, $MSE = 113.87$, $p < .01$; $F2(1, 58) = 2.37$, $MSE = 158.75$, $p > .1$), where participants had significantly higher accuracy scores in the VA than in the AA condition ($M1$: VA = 78.52 %, AA = 73.50 %). Figure 7.1 provides a summary of the findings.

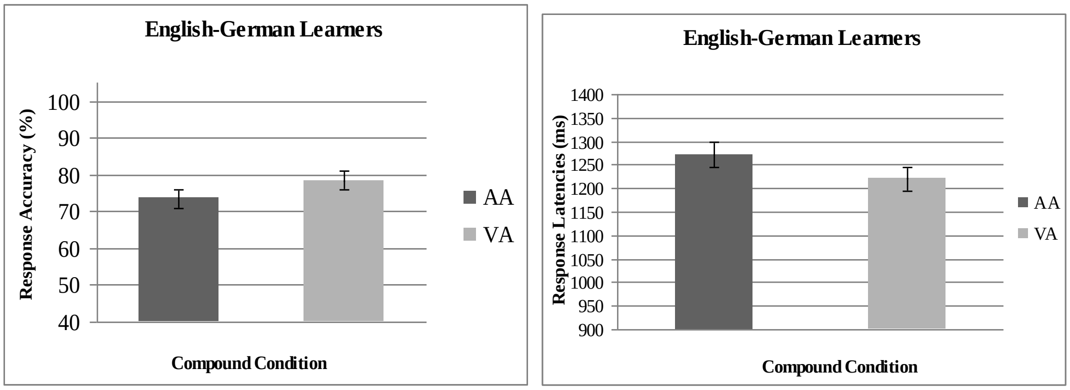
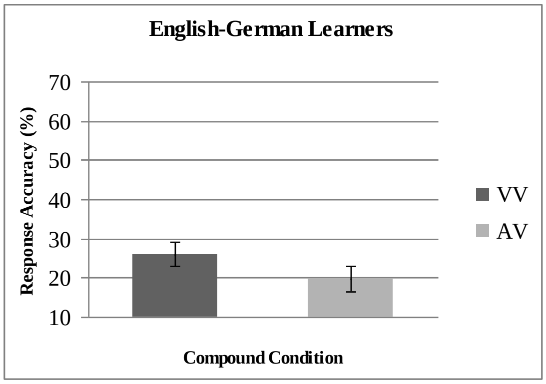


Figure 7.1: Mean translation accuracy and mean response latencies for English-German learners as a function of compound condition where AA = agreement, agreement (i.e., the gender of the compound agrees with both constituents) and VA = violation, agreement (i.e., the gender of the compound agrees with only the second constituent).

7.4.2.2 No Trials

Analyses of the "no" trials revealed no effect of compound type (Fs < 1) with equivalent response latencies in the agreement violation condition, AV, and the double violation condition, VV ($M1$: AV = 1562 ms, VV = 1557 ms). In the accuracy analysis, a main effect of compound type was found ($F1(1, 81) = 22.24$, $MSE = 74.59$, $p < .001$; $F2(1, 58) = 5.16$, $MSE = 115.65$, $p < .05$), with higher accuracy scores in the VV than the AV condition ($M1$: VV = 26.04 %, AV = 19.68 %). Figure 7.2 provides a summary of the findings.



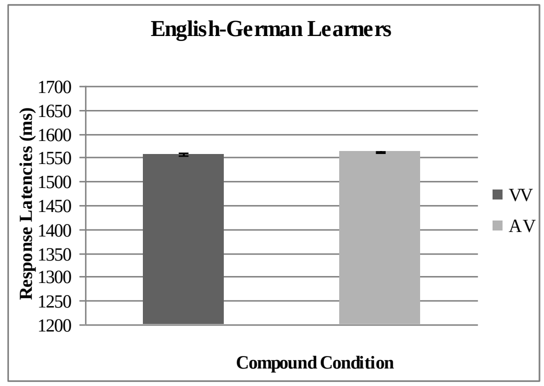


Figure 7.2: Mean translation accuracy and mean response latencies for English-German learners as a function of compound condition where VV = violation, violation (i.e., the gender of the compound violates both constituents) and AV = agreement, violation (i.e., the gender of the compound violates only the second constituent).

7.4.2.3 Proficiency in Compounding

Next, the role of proficiency in modulating these effect was investigated. Unlike in Experiment 1A, participants could not be divided according to L2 proficiency based on their accuracy in the gender assignment task because investigation of the homogeneity of variances using Levene's test indicated that variances between groups were significantly different from each other ($F(1, 80) = 16.22$, $p < .001$ for the AA condition, and $F(1, 80) = 6.38$, $p < .05$ for the AV condition). Attempts to transform the data using the natural log, inverse square root, reciprocal, square root, square, or cube failed to equalize the variance.

Instead, a difference score was computed for response latencies between the AA and VA condition for the "yes" trials (AA – VA), and the AV and VV condition for the "no" trials (AV-VV), and correlations with participants' gender assignment scores were performed. For the "yes" trials, differences between the two measures appear to be positively related to participants' scores on the gender assignment task, but this correlation was not significant ($r = .067$, $p > .1$). For the "no" trials, differences between the two measures appear to be negatively correlated to participants' scores on the gender assignment task, but this correlation was again not significant ($r = -.13$, $p > .1$). These preliminary analyses suggest that proficiency may not modulate sensitivity to gender in compounding.

The results of these analyses suggest that English-German L2 learners may in fact be sensitive to gender in compounding to a limited degree. For correct translation trials, participants showed significantly faster and more accurate responses to items in the VA than the AA condition. Conversely, for the incorrect translation trials, participants were more accurate to reject compounds in the VV than the AV condition. Proficiency, however, did not appear to modulate these effects. In some respects, these findings are puzzling, because the direction of the effects in the correct translation trials goes against the predictions, in which faster responses were predicted for the AA condition. One possibility is that participants simply knew the compounds in the VA condition better, but this would not explain the reversal of the effect for the incorrect translations, in which one would then predict faster rejection of items with which participants were more familiar, in this case the AV condition. These unexpected results will be discussed in greater detail in the general discussion.

Similar to Experiment 1B, in a next step, the paradigm was used with a group of German-English L2 learners in order to further investigate compounding. Like the caveat in Experiment 1A and 1B, direct comparisons between language groups are difficult to make since the English-German group performed forward translation, but the German-English group performed backward translation. Results of the study, however, can still provide insight into how gender processing may proceed in compounding and as a result may shed light on the previous results of the English-German bilinguals.

7.5 Experiment 3B

7.5.1 Method

7.5.1.1 Participants

The same 35 participants as in Experiment 1B participated in the present experiment.

7.5.1.2 Materials

The same materials as in Experiment 3A were used.

7.5.1.3 Procedure

The same procedure as in Experiments 1A, 1B, and 3A was followed.

7.5.2 Results and Discussion

7.5.2.1 Yes Trials

Among the German-English learners, analyses of the "yes" trials showed no main effect of compound type ($F1(1, 34) = 1.67$, $MSE = 4490.23$, $p > .1$; $F2 < 1$**;** $M1$: VA = 967 ms, AA = 988 ms). Analyses for arcsin transformed accuracy data revealed a main effect of compound type ($F1(1, 34) = 34.31$, $MSE = 64.89$, $p < .001$; $F2(1, 58) = 5.89$, $MSE = 170.56$, $p < .05$), where participants had significantly higher accuracy scores in the VA than in the AA condition (uncorrected $M1$: VA = 94.86 %, AA = 86.10 %, $M2$: VA = 94.89 %, AA = 86.13 %). Figure 7.3 provides a summary of the findings.

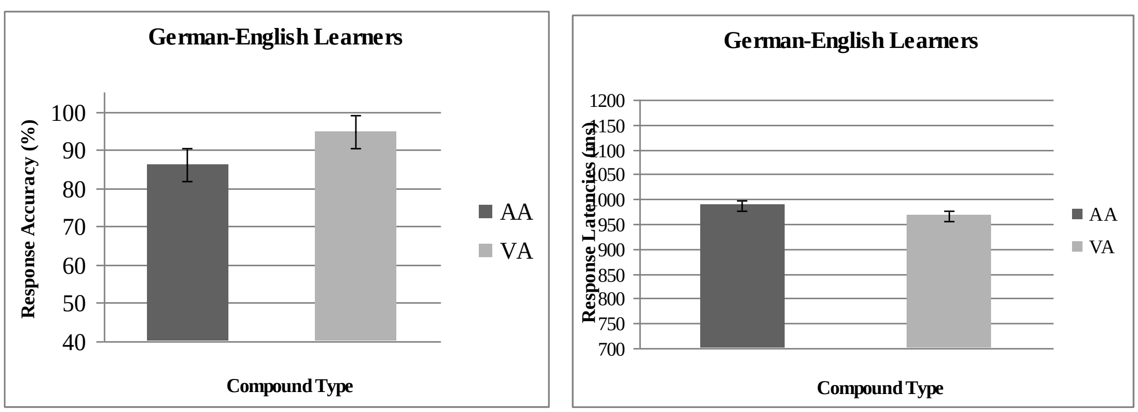
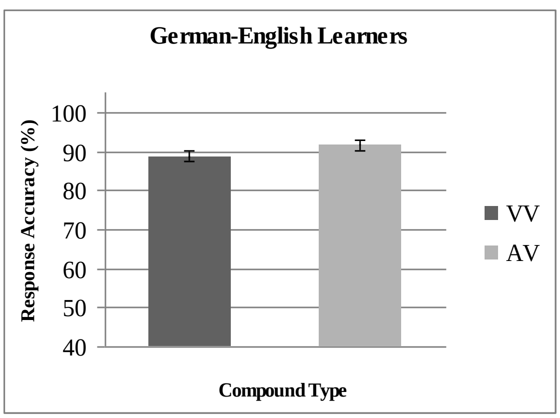


Figure 7.3: Mean translation accuracy and mean response latencies for English-German learners as a function of compound condition where AA = agreement, agreement (i.e., the gender of the compound agrees with both constituents) and VA = violation, agreement (i.e., the gender of the compound agrees with only the second constituent).

7.5.2.2 No Trials

Analyses of the "no" trials revealed a significant effect of compound type in the participant analysis ($F1(1, 34) = 15.88$, $MSE = 3982.00$, $p < .001$; $F2(1, 58) = 3.21$, $MSE = 16866.78$, $p < .1$), with faster response latencies in the agreement violation condition, AV, than in the double violation condition, VV (*M*1: AV = 951 ms, VV = 1011 ms). Analyses for arcsin transformed accuracy data revealed a main effect of compound type in the participant analysis ($F1(1, 34) = 5.48$, $MSE = 51.44$, $p < .05$; $F2 < 1$), with higher accuracy scores in the AV than the VV condition (uncorrected *M*1: VV = 88.76 %, AV = 91.62 %). Figure 7.4 provides a summary of the findings.



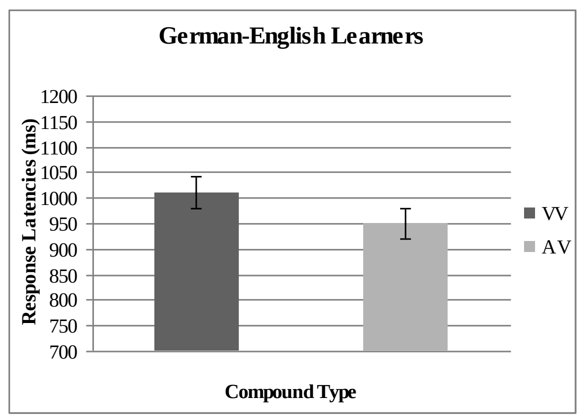


Figure 7.4: Mean translation accuracy and mean response latencies for English-German learners as a function of compound condition where VV = violation, violation (i.e., the gender of the compound disagrees with both constituents) and AV = agreement, violation (i.e., the gender of the compound agrees with only the second constituent).

7.5.2.3 Proficiency in Compounding

To explore the possible role of L2 proficiency in modulating these translation effects, a difference score was again computed for the "yes" (VA-AA) and the "no" trials (VV-AV) and correlated with L2 self-rating scores, since the scores on the gender assignment task in the participants' L1 would not reflect sensitivity to L2 grammatical knowledge. Results revealed a negative correlation between latency differences and L2 self-ratings, but this correlation was not significant (r = -.086, p > .1). Likewise, the positive correlation between differences on the "no" trials and L2 self-ratings was not significant (r = .187, p > .1). These results suggest no modulation of gender sensitivity in compounding by L2 proficiency.

Similar to the English-German L2 learner group, there seems to be partial support for gender sensitivity in compounding within the translation recognition paradigm. Accuracy data for the "yes" trials revealed significantly higher accuracy rates in judging translations in the VA condition than the AA condition. Data for the "no" trials revealed both faster response latencies and higher accuracy scores for the AV condition than the VV condition. This sensitivity, however, does not seem to be modulated by L2 proficiency.

7.6 General Discussion

Across both language groups, there is evidence that participants processed compounds in each of the conditions differently. One difficulty in interpreting the results, especially for English-German L2 learners, is the inability to control for individual learner knowledge of compounds. Since, by virtue of the stimuli, different items had to be used in each condition, it is plausible that compounds in one condition may have been better known to learners than compounds in the other condition. While the items were matched on native speaker norms, they were not matched for norms based on the present English-German L2 sample, which may have lead to between-condition variance. Future work would need to account for this individual learner variance in order to address the contribution of compound type to processing differences.

While the pattern of data between language groups is not identical, it is striking that both the English-German and the German-English L2 learners show differences in the VA over the AA condition for “yes” responses. While English-German L2 learners showed faster response latencies and higher accuracy scores in the VA condition, German-English L2 learners showed higher accuracy scores, but no latency differences. Assuming that item variability is not the source of these effects, this finding may point to the fact that these two language groups process these types of compounds similarly. Further, these results suggest that both groups may be sensitive to the internal gender violation/agreement between constituents and by extension may point to decomposition of the morphemes. If participants had been processing the two types of compounds as whole units, one would have predicted no differential pattern across conditions. The lack of an effect in the reaction time data of the German-English group, however, sheds some doubt on this interpretation, as one would have expected a differential cost in processing as well. And yet, German-English L2 learners also show a dissociation between the groups in both response latency and response accuracy for the “no” trials. The English-German group only shows this effect for the “no” trials in the response accuracy, and in the opposite direction (higher accuracy in the VV than the AV condition). The fact that results are less clear for the English-German group for the "no" trials is less surprising, given the already established fact from Experiment 1A that rejecting translations based on their gender is a difficult task. If anything, response accuracies are even lower in the present experiment (between 20 and 25 % as opposed to 30 or 35 %).

A particularly relevant factor for the present results is the fact that participants in both language groups are engaging in two tasks, one semantic, and the other grammatical, as they complete the translation recognition task. For the English-German L2 learners, the semantic task of matching the content of the nouns across translations might be easier relative to the task of judging the gender accuracy for a given translation. For the German-English L2 learners, the reverse should be true, where judging the gender accuracy is easy (since this is based on their knowledge of their L1), but judging the semantic match of the two words could be more difficult. These differences may help explain the differences both across "yes" and "no" trials, and across language groups. One way to dissociate the semantic and grammatical effects would be to use a different paradigm which does not require the use of both types of judgment. A task similar to Hofmann (2006) might provide such a paradigm, in which participants are shown an adjective such as *feminine*, *masculine*, or *neuter* which is followed by a German word. Participants then decide whether the pairing is correct or not. To approximate the present paradigm, participants could be shown the actual article followed by the noun (DER – SPIELPLATZ) and then decide whether the pairing is correct or not. A traditional translation recognition paradigm could be used to investigate semantic effects, in which participants would judge translation presentation such as PLAYGROUND –

SPIELPLATZ. By completing additional tasks such as these, it may be possible to begin disentangling the independent contributions of semantic and grammatical processing in translation recognition.

Chapter 8

General Discussion

In the past decade there has been a notable increase in research investigating bilingualism and second language (L2) learning. Many recent studies suggest that the cognitive architecture underlying L2 performance is remarkably adaptive to accommodate acquisition of a new language. However, some limitations remain. The goal of the present study was to further clarify constraints to language learning and help address questions about L2 learning that have not yet been fully resolved. The experiments examined the degree to which L2 learners and proficient bilinguals are able to fully access grammatical and morphological features of the L2. The specific aim of the study was to identify the ability of intermediate and advanced English-German bilinguals to comprehend the assignment of grammatical gender and to interpret the meaning of compounds. Grammatical gender is a feature that is typically considered difficult to acquire in the L2. Particularly for those whose native language does not mark gender, such as English, the question has been raised whether full acquisition of gender can take place and under which circumstances. The current study aimed to contribute to the resolution of current debates about the degree to which L2 learners can acquire and process subtle aspects of the non-native language. In this chapter, the major findings of each of the experiments will first be summarized. Results will then be interpreted within the context of the developmental and representational issues raised in Chapters 1 and 6.

After addressing some additional considerations, general conclusions will be made and further directions will be proposed.

8.1 Summary of Findings

Experiment 1A set out to investigate the sensitivity of L2 learners of German to grammatical gender and introduced the paradigm of translation recognition with simple nouns as a way to investigate gender processing in L2 learners. An additional goal of the experiment was to investigate the role of proficiency in modulating sensitivity to gender in late L2 learners. Results indicated that participants had particular difficulties in rejecting correct noun translations with the wrong gender. An increase in proficiency did not change how gender was being processed: While performance improved overall with increasing proficiency, even the more advanced learners continued to show difficulties in judging the wrong article, right noun condition. With increased proficiency, they also did not show sensitivity to the gender of wrong translations when it matched the gender of the correct translation. In contrast, results from Experiment 1B with German-English L2 learners showed robust gender effects, in which participants took longer to reject wrong translations whose gender matched the gender of the correct translation compared to translations whose gender did not match that of the correct translation. Results suggest that native speakers of German are sensitive to gender matches and mismatches across translations, and leave open the possibility that L2 learners of German who achieve native-like language competency may eventually begin to show sensitivity to gender using this task.

In a next step, data from the gender assignment task were analyzed by choosing words from the critical items which corresponded to three types of gender assignment patterns: typical, in which nouns ending in *–e* are typically assigned the feminine gender; ambiguous, in which nouns ending in a consonant are either assigned the masculine or neuter gender; and atypical, in which nouns ending in a consonant are assigned the feminine gender or nouns ending in *–e* are assigned the masculine or neuter gender. Results suggested that both English-German L2 learners and German-English L2 learners were sensitive to this gender distribution. Both groups showed slower response times and decreased accuracy in assigning gender to nouns with ambiguous gender assignment patterns. English-German L2 learners, however, showed an additional dissociation between the three levels. While German-English learners had similar response times and accuracy rates on the typical and atypical categories, English-German learners showed significant differences between all three categories, with fastest response rates in the typical category, followed by the atypical category and then the ambiguous category. Analyses by proficiency group showed that this dissociation between all three levels held only for higher proficiency English-German learners, suggesting, unlike the proficiency results of Experiment 1A, that proficiency modulates sensitivity to gender in a task that is more under the participant's control. These results suggest that with increased proficiency, L2 learners of German become more sensitive to the phonological distributions in the L2, which may be a way that they are able to behaviorally approximate native-like gender use.

In Experiment 2, ERPs were used with the same paradigm and similar materials from Experiment 1 to investigate the possibility that the timing parameters of Experiment

1 may not have been sensitive enough to detect on-line gender sensitivity. Preliminary data from Experiment 2 statistically showed no sensitivity to gender in noun processing in the sample of 9 participants and underscored the sensitivity of L2 learners of German to semantics in translation. Together with the results from Experiment 1, the statistical results from Experiment 2 suggest that more on-line measures are able to distinguish L2 performance from native performance, showing an inability to exploit gender in on-line processing. Visual inspection of the components, however, indicated a reduced amplitude for both the N250 and the N400 in the wrong article, right noun condition and a further reduction in amplitude for the right article, right noun condition for both of these two components. Had the effect been purely semantic, one would have expected these two components to show similar amplitudes. There is some indication therefore that there may be the beginnings of a sensitivity to gender in this sample of participants. Further testing with a larger sample is needed to tease apart these effects.

In the final experiment, L2 gender processing was used as a way to investigate how morphologically complex words are represented in the L2 mental lexicon. The same paradigm as in Experiments 1 and 2 was used, but this time with German compound words. Results for the English-German L2 group in Experiment 3A revealed sensitivity in processing internal gender agreement in compounds for both “yes” trials as well as “no” trials, although the pattern of data were not in the predicted direction. Proficiency, however, did not appear to modulate these effects. German-English L2 participants in Experiment 3B also showed sensitivity to gender in compounding for both “yes” and “no” trials, but again not in the predicted direction. Proficiency also did not modulate these effects.

Together, the results of these experiments confirm previous results on the difficulty of L2 gender processing in German (e.g., Sabourin, Stowe, & de Haan, 2006) and also appear to show dissociations between tasks that require more automatic processing (Experiments 1 and 2) and those that are under the participant's control (overt gender assignment). While clear developmental patterns exist in the data, showing faster and more accurate performance with increasing proficiency, increasing proficiency itself does not appear to be sufficient to modulate these effects, nor confer native-like processing of gender. Overall, L2 learners of German in this sample of participants remained relatively insensitive to grammatical gender in on-line processing, although there are hints in the ERP data, as well as Experiment 3, of an emerging sensitivity to gender. The analysis of the gender assignment data also suggests that L2 learners may rely on distributional properties of gender to bootstrap their way into the gender system, displaying a more native-like behavior (distinct from native-like processing) on typically-marked nouns.

8.2 Implications for Late L2 Learning

8.2.1 The Critical Period Hypothesis

In Chapter 1, several models of L2 learning were introduced which explicitly or implicitly address the claims of the Critical Period Hypothesis. In the Failed Functional Feature Hypothesis (Hawkins & Chan, 1997), the ability to learn language features becomes difficult to impossible after a critical period. The Full Transfer/Full Access

Hypothesis (Schwartz & Sprouse, 1996), on the other hand, proposes that L2 learners can develop language structures unique to the L2 even after a critical period. In the Competence Deficit Approach (Jiang, 2000), incomplete lexical representations of L2 specific features hamper automatic, but not explicit, processing. Computational models (Bates & MacWhinney, 1981) and statistical learning models (Saffran, Aslin, & Newport, 1996) were also considered, which draw on shared cognitive learning mechanisms and linguistic cues and distributional probabilities that drive language learning.

The evidence as provided by the current study suggests constraints on the extent to which this sample of late L2 learners had learned grammatical gender at the point in language development at which they were tested. Taking these constraints into consideration, within the possibilities framed by the theories above, three general explanations of the data are possible:

First, L2 learners are not acquiring the structure, either on-line or off-line, which would provide evidence for a hard constraint on the late acquisition of grammatical structures not present in the L1. The fact that English-German participants are showing some sensitivity to gender would suggest that a strong version of a hard constraint (i.e., Hawkins & Chan, 1997) is not tenable.

A second possibility is that L2 learners are acquiring the structure but using a qualitatively different computing structure. This possibility would be similar to the Competence Deficit Approach, suggesting difficulties in on-line processing, but the ability to use strategies and rules for more successful explicit processing. In fact, this possibility concurs with the data presented in the current study, which showed particular difficulties for L2 learners of German in Experiments 1 and 2, but revealed sensitivity to gender rules and distributional properties in the analysis of the gender assignment task. In the less speeded and more explicitly gender-focused gender assignment task, participants showed a sensitivity to gender for a subset of nouns with a particularly salient phonological marking, and sensitivity to these cues increased with increasing proficiency.

A third possibility is that late L2 learners may approximate the behavior of native speakers, but not have the cognitive resources to respond quickly enough. As a result, this slower time course leaves the system more open to other effects. The seeming salience of the semantics for L2 learners in the translation recognition task documented here and elsewhere (Sunderman & Kroll, 2006) seem particularly relevant for this point. It is possible that a decrease in the availability of cognitive resources and the resulting slower processing speed of the English-German L2 learners left the system more open to semantic effects. Significantly, the German-English L2 group was still sensitive to gender effects despite the semantic nature of the task, suggesting their ability to deemphasize semantic information in performing the translation task. Particularly relevant to this third possibility is research such as that by McDonald (2006), who showed that native speakers, when put under cognitive stress or load, actually perform like non-native speakers. Similar to the pattern seen in L2 learners, these native speakers showed particular difficulties with some structure such as the regular past, while other structures such as word order seemed less susceptible to stress.

While the pattern of data presented in this study cannot fully distinguish between the latter two alternatives, there does seem to be enough evidence to reject the strong version of the Failed Functional Feature Hypothesis which constrains acquisition of L2 specific features to a critical period.

8.2.2 Representation of Gender in the L2 Lexicon

Chapter 2 also addressed several monolingual models of gender representation in the mental lexicon. To review briefly, in the Intralexical Model (Fodor, 1983; Swinney, 1979), the grammatical gender of a noun is part of its lexical representation. Gender information becomes activated with the appearance of the gendered article and can therefore show prelexical or lexical effects with the activation of the noun. Interactive Models (Bates, Devescovi, Hernandez, & Pizzamiglio, 1996) are most consistent with computational learning models and conceive of a nonmodular view of the lexicon in which semantic and syntactic context interact with lexical processing. In this account, gender can serve to constrain lexical access, and, similar to the Intralexical Model, also predicts prelexical and lexical effects. Prelexical effects assume that participants are predicting lexical candidates based on the available contextual semantic and syntactic information in order to narrow the possible selection. Seeing a gendered article would thus preactivate lexical items of the same gender (for a more detailed description of this process, see Friederici & Jacobson, 1999).

These two theories can be distinguished from the Checking Model by Friederici and Jacobson (1999), whereby gender becomes activated with the appearance of the gendered article, but is only used in a post-lexical checking mechanism after activation of the noun. Recall that previous research has questioned the viability of the Checking Model for L2 gender processing (Taraban & Kempe, 1999) on the basis of the difficulty L2 learners display in learning gender in the first place.

The validity of these theories has been largely tested in lexical decision tasks and gender priming studies. In the present study, due to the nature of the task, an additional variable involving semantic priming becomes relevant. Assume that a participant sees the translation pairs THE DOG (der Hund) – DER TISCH (the table). Upon seeing the word *dog*, the translation equivalent *Hund* should become activated. In a postlexical checking view, the article *der* does not become relevant until after the German noun, *Tisch*, becomes activated, at which point the translation should have been rejected. The article should bear no effect on the translation decision. According to a prelexical account, the article would send activation to other masculine (der) words, including the word *Hund*. The word *Hund* has now received activation from the English noun and the German article, and this additional activation must be overcome for the participant to reject the wrong translation. The prelexical account, but not the postlexical account, would predict different processing between this condition and a condition in which the gender does not match the gender of the translation (e.g., THE DOG – DIE PUPPE (the doll)). It would seem that a prelexical account could account best for the German-English L2 speaker results with longer latencies in the right article, wrong noun condition as illustrated above. The fact that English-German L2 speakers do not show this effect could suggest that their lexical representations are incomplete (Jiang, 2000), so that seeing a gendered article does not spread activation to the relevant noun.

8.2.3 Morphological Access

As discussed in Chapter 6, an important question in both the monolingual and bilingual domain that has not yet been fully resolved is how morphologically complex words are represented in the mind. While some theories claim morphologically complex words are represented as whole words (Butterworth, 1983), others have argued for decomposition of words (Taft, 2004) or even a race model between these two possibilities (Gunter & Friederici, 2003). Research reviewed in the bilingual domain is no less clear on how the bilingual mind processes words. On one hand, there seems to be evidence for decomposition of L2 words (e.g., Lehtonen & Laine, 2003), while on the other, a default to treat morphologically complex L2 words as whole units (Silva & Clahsen, 2008). The results of Experiment 3 on gender processing in German compound nouns provide some preliminary evidence to these questions, suggesting the possibility of decomposition in both English-German and German-English L2 learners.

One limitation in Experiment 3 was that by virtue of testing a broad range of L2 proficiencies, materials had to be relevant to lower proficiency participants. As a result, items could only be selected from a small pool of possible compounds, and item characteristics such as frequency and length, while controlled, could not be orthogonally manipulated. Past research as reviewed in Chapter 5 has pointed to the relevance of these factors in bilingual decomposition (e.g., Portin, Lehtonen, & Laine, 2006), and future research would need to attempt to account for their influence on L2 processing of compounds.

Another limitation in the present investigation is that the results of Experiment 3 in many ways depend on the presupposition of gender sensitivity in late L2 learners. As the results of Experiment 1 and 2 illustrated, and as discussed above, this assumption may have not been met in this sample of English-German L2 learners. However, while alternate explanations of the data are possible, there does seem to be some evidence in both language groups for decomposition.

8.3 Additional Considerations

An important additional consideration introduced briefly in this chapter and in previous chapters is the extent to which the nature of the task biases participants towards semantic processing. On the one hand, participants might not be able to access grammatical structures within the timeframe of the experiment, particularly if it is a later lexical process (Friederici & Jacobson, 1999). On the other hand, the task itself might minimize the influence of grammatical processing, making the reduced gender sensitivity in the present experiments less surprising. By focusing on translation which typically emphasizes semantic matches or mismatches between words, gender processing may not have come to bear on the translation decision being made.

Another possibility is that the language environment plays a role in modulating gender sensitivity. It is possible that, with enough context, sensitivity to gender in the English-German learner group would begin to emerge. Research such as Schwartz and Kroll (2006) looking at bilingual language processing in sentences suggests that the sentence context only exerts an influence under high constraint conditions, in which rich semantic cues were available (see also Van Hell, 1998; Duyck, Van Assche, Drieghe, & Hartsuiker, 2007). The fact that many studies investigating gender have looked at these effects in sentence contexts offers the possibility that the broader linguistic framework around the noun phrase may in fact be relevant.

The more global context around the L2 may also be important in influencing gender processing. Research by Linck, Sunderman and Kroll (in preparation) showed that immersion context affects the ability to suppress the L1, while findings by Elston-Güttler, Gunter, and Kotz (2005) suggest that the language environment influences priming effects in sentence processing. In the study by Elston-Güttler and colleagues, by preceding the all-L2 task with a movie in either the L2 or the L1, cross-language priming effects were either absent (after watching the L2 movie) or enhanced (after watching the L1 movie) leading the researchers to propose that the preceding L2 exposure allowed participants to “zoom into” the L2 thereby decreasing cross-language effects. A follow-up study to the current experiments which could address the idea of more global context effects would be to test English-German participants immersed in the L2 environment. Inducing the zoomed-into L2 mode may be critical in developing sensitivity to subtle grammatical features of the L2. Analagous perhaps to a developmental stage, zooming may serve a similar purpose as the retrieval of the L1 translation equivalent in mediating the L2, as proposed in the Revised Hierarchical Model (Kroll & Stewart, 1994). In the proposal of the RHM, the L1 provides a temporary scaffold until the L2 learner can directly mediate the L2. Similarly, immersion contexts may serve as a stepping stone in developing grammatical sensitivity.

8.4 Conclusions and Future Directions

As indicated earlier, the merging of semantic and lexical processing in the current paradigm makes it difficult to tease apart the independent contributions of each, although investigating gender processing in a context that merges semantic and lexical processing was at the same time an important innovation of the current study. An important next step would be to investigate the individual influence of each using an adapted, or different, paradigm. As proposed at the end of Experiment 3, a typical translation recognition task with bare noun translations in conjunction with a gender matching task may begin to address this question.

The current study thus provides more evidence to the growing body of literature documenting constraints to L2 learning. Particularly the lack of modulation of effects by proficiency points to the extreme difficulty in learning certain L2-specific language structures. By investigating a structure particularly difficult to acquire, however, the approach to ultimate attainment was conservative, leaning if anything towards a null result in findings. In this respect, the glimpses of gender sensitivity in both the gender assignment task, in trends of the ERP data, and in the compound study are particularly hopeful, suggesting that perhaps with a broader sample of proficiency, or more participants with a higher L2 proficiency, and with possible methodological adaptations, gender sensitivity may begin to emerge.

Bibliography

Abutalebi, J., & Green, D. (2007). Bilingual language production: The neurocognition of language representation and control. *Journal of Neurolinguistics*, *20*, 3, 242-275.

Abutalebi, J., Cappa, S.F., & Perani, D. (2005). What can functional neuroimaging tell us about the bilingual brain? In Kroll, J.F., & De Groot, A.M.B. (Eds). *Handbook of Bilingualism: Psycholinguistic Approaches*. New York: Oxford University Press.

Akhutina, T., Kurgansky, A., Polinsky, M., & Bates, E. (1999). Processing of grammatical gender in three-gender system: Experimental evidence from Russian. *Journal of Psycholinguistic Research, 28,* 695-713.

Altarriba, J. (1990). Constraints of interlingual facilitation effects in priming in Spanish-English bilinguals. *Unpublished dissertation*, Vanderbilt University Nashville.

Andrews, S., Miller, B., & Rayner, K. (2004). Eye movements and morphological segmentation of compound words: There is a mouse in mousetrap. *European Journal of Cognitive Psychology, 16,* 285-311.

Ardal, S., Donald, M. W., Meuter, R. Muldrew, S., & Luce, M. (1990). Brain responses to semantic incongruity in bilinguals. *Brain and Language*, 39, 187–205.

Aronoff, M. (1976).*Word formation in generative grammar*. Linguistic Inquiry Monograph 1. Cambridge, MA: MIT Press.

Au, T. K., Knightly, L. M., Jun, S.-A., & Oh, J. S. (2002). Overhearing a language during childhood. Psychological Science, 13, 238-243. *Reprinted in: Robinson-Riegler,

B. & Robinson- Riegler, G. L. (Eds.) (2003). Readings in Cognitive Psychology: Applications, Connections, and Individual Differences. Boston: Allyn and Bacon.

Au, T. K., Oh, J., Knightly, L. M., Jun, S., & Romo, L. (2008). Salvaging a Childhood Language. *Journal of Memory and Language, 58,* 998-1011.

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). *The Celex Lexical Database.* Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania.

Baker, C. & Prys Jones, S. (Eds.) (1998). *Encyclopedia of Bilingualism and Bilingual Education*. Clevedon: Multilingual Matters.

Barber, H., & Carreiras., M. (2005). Grammatical Gender and Number Agreement in Spanish: An ERP Comparison. *Journal of Cognitive Neuroscience, 17,*137-153.

Bates, E., & MacWhinney, B. (1981). Second language acquisition from a functionalist perspective: Pragmatic, semantic and perceptual strategies. In H. Winitz (Ed.), *Annals of the New York Academy of Sciences conference on native and foreign language acquisition*. New York: New York Academy of Sciences.

Bates, E., & MacWhinney, B. (1989). Functionalism and the Competition Model. In B. MacWhinney, & E. Bates (Eds.), *The crosslinguistic study of sentence processing*. New York: Cambridge University Press.

Bates, E., Devescovi, A., Hernandez, A., & Pizzamiglio, L. (1996). Gender priming in Italian. *Perception & Psychophysics, 85*(7), 992-1004.

Bates, E., Devescovi, A., Pizzamiglio, L., D'Amico, S., & Hernandez, A. (1995). Gender and lexical access in Italian. Perception & Psychophysics, 57(6), 847-862.

Bates, E., Elman, J., & Li, P. (1994). Language in, on and about time. In M. Haith, J. Benson, R. Roberts & B. Pennington, (Eds.), *The development of future-oriented processes*. Chicago: The University of Chicago Press, 293-321.

Bialystok, E. & Hakuta, K. (1994). In other words: The science and psychology of second language acquisition. New York: Basic Books.

Bialystok, E. (2001). Metalinguistic aspects of bilingual processing. *Annual Review of Applied Linguistics, 21*,169-181.

Bialystok, E. (2002). On the reliability of robustness: A reply to DeKeyser. *Studies in Second Language Acquisition, 24*, 481-488.

Bialystok, E., & Miller, B. (1999). The problem of age in second language acquisition: Influences from language, task, and structure. *Bilingualism: Language and Cognition, 2*, 127-145.

Bialystok, E., Craik, F. I. M., Grady, C., Chau, W., Ishii, R., Gunji, A., & Pantev, C. (2005). Effect of bilingualism on cognitive control in the Simon task: Evidence from MEG. *NeuroImage, 24*, 40-49.

Bialystok, E., Craik, F. I. M., Klein, R., & Viswanathan, M. (2004). Bilingualism, aging, and cognitive control: Evidence from the Simon task. *Psychology and Aging, 19*, 290-303.

Billman, D. (1989). Systems of correlations in rule and category learning: Use of structured input in learning syntactic categories. *Language and Cognitive Processes, 4*, 127-155.

Birdsong, D. (1992). Ultimate attainment in second language acquisition. *Language, 68*, 706–755.

Birdsong, D. (2005). Nativelikeness and non-nativelikeness in L2A research. *International Review of Applied Linguistics, 43*, 319-328.

Birdsong, D., & Flege, J. E. (2001). Regular-irregular dissocations in the acquisition of English as a second language. In A. H.-J. Do, L. Dominguez, & A. Johansen (Eds.), *BUCLD 25: Proceedings of the 25th Annual Boston University Conference on Language Development* (pp. 123-132). Boston: Cascadilla Press.

Birdsong, D., & Molis, M. (2001). On the evidence for maturational constraints in second-language acquisition. *Journal of Memory and Language, 44*, 1-15.

Bongaerts, T. (1999). Ultimate attainment in foreign language pronunciation: The case of very advanced late foreign language learners. In D. Birdsong (Ed.), *Second language acquisition and the critical period hypothesis* (pp. 133-159). Mahwah, NJ: Erlbaum.

Bordag, D. & Pechmann, T. (2008). Grammatical gender in speech production: Evidence from Czech. *Journal of Psycholinguistic Research*, *37*(2), 69-85.

Bordag, D. and Pechmann, T. (2007). Factors influencing L2 gender processing. *Bilingualism: Language and Cognition, 10*, 299-314.

Bordag, D., Opitz, A. & Pechmann, T. (2006) Gender processing in first and second langauges: The role of noun termination. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*, 1090-1101.

Bosch, L. & Sebastian-Galles, N. (1997). The role of prosody in infants' native-language discrimination abilities: the case of two phonologically close languages, In *EUROSPEECH-1997*, 231-234.

Bosch, L.; Sebastián-Gallés, N. (2003). Simultaneous bilingualism and the perception of a language-specific vowel contrast in the first year of life. *Language & Speech*, *46*, 217-243.

Brenders, P., van Hell, J., & Dijkstra, T. (2007). The development of lexical and conceptual processing in classroom L2 learners: ERP and behavioral evidence. Paper presented at the meeting: International Symposium on Bilingualism (ISB) 6, Hamburg, Germany.

Brooks, P. J., Braine, M. D. S., Catalano, L., Brody, R. E., & Sudhalter, V. (1993). Acquisition of gender-like noun subclasses in an artificial language: the contribution of phonological markers to learning. *Journal of Memory and Language, 32,* 76-95.

Brown, C. & Hagoort, P. (1993). The processing nature of the N400: Evidence from Masked priming. *Journal of Cognitive Neuroscience 5*, 34-44.

Bruhn de Garavito, J. (1999). The SE constructions in Spanish and near-native competence. *Spanish Applied Linguistics. 3,* 247-295.

Brysbaert, M., & Mitchell, D. C. (2000). The failure to use gender information in parsing: A comment on van Berkum, Brown, and Hagoort (1999). *Journal of Psycholinguistic Research, 29*(5), 453-465.

Brysbaert, M., & Mitchell, D. C. (1996). Modifier attachment in sentence parsing: Evidence from Dutch. *Quarterly Journal of Experimental Psychology, 49A*, 664-695.

Butterworth, B. (1983). Lexical representation. In B. Butterworth (ed.), *Language production (Vol. II): Development, Writing and Other Language Processes*. London: Academic Press. 257-294.

Bybee, J. (1995). Regular morphology and the lexicon. Language and Cognitive Processes, 10, 425-455.

Caramazza, A., Laudanna, A., & Romani, C. (1988). Lexical access and inflectional morphology. *Cognition, 28*, 297-332.

Carroll, S. (1989). Second-language acquisition and the computational paradigm. *Language Learning*, 39, 535-594.

Chomsky, N. (1980). *Rules and Representations*. New York: Columbia University Press.

Clahsen, H. & C. Felser 2006. Grammatical processing in language learners. *Applied Psycholinguistics, 27*, 3-42.

Clahsen, H., Penke, M., & Parodi, T. (1993). Functional categories in early child German. *Language Acquisition* 3, 395-429.

Coles M. G. H, & Rugg M. D. (1995). Event-related brain potentials: an introduction. In MD Rugg, MGH Coles (Eds.), Electrophysiology of Mind: Event-Related Brain Potentials and Cognition. Oxford University Press.

Coltheart, M. (1978). Lexical access in simple reading tasks. In G. Underwood, Ed., *Strategies of Information Processing.* London: Academic Press.

Coltheart, M. (1981). The MRC Psycholinguistic Database. *Quarterly Journal of Experimental Psychology, 33A*, 497-505.

Coppetiers, R. (1987). Competence differences between native and non-native speakers. *Language*, 63, 544-573.

Costa, A., Miozzo, M., & Caramazza, A. (1999). Lexical selection in bilinguals: Do words in the bilingual's two lexicons compete for selection? *Journal of Memory and Language*, *41*, 365-397.

Coulson, S., King, J. W., & Kutas, M. (1998). Expect the unexpected: Event-related brain response to morphosyntactic violations. *Language and Cognitive Processes,13*(1), 21-58.

Crutch S. J. (2006). Qualitatively different semantic representations for abstract and concrete words: Further evidence from the semantic reading errors of deep dyslexic patients. *Neurocase, 12,* 91-97.

Crutch, S. J., & Warrington, E. K. (2005). Abstract and concrete concepts have structurally different representational frameworks. *Brain, 128,* 615-627.

De Groot, A. M. B. (1992). Determinants of word translation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18,* 1001-1018.

De Groot, A. M. B., & Comijs, H. (1995). Translation recognition and translation production: Comparing a new and an old tool in the study of bilingualism. *Language Learning, 45,* 467-509.

De Groot, A. M. B., Delmaar, P., & Lupker, S. L. (2000). The processing of interlexical homographs in translation recognition and lexical decision: Support for nonselective access to bilingual memory. *Quarterly Journal of Experimental Psychology, 52a*, 397-428.

de Groot, A. M. B. (1995). Determinants of bilingual lexicosemantic organization. *Computer Assisted Language Learning, 8*, 151-180.

DeKeyser, R. M. (2000). The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition, 22,* 499-533. (2000).

DeKeyser, R. M., Larson-Hall, J. (2005). "What Does the Critical Period Really Mean?" In J. F. Kroll and A. M. B. de Groot (Eds.) *Handbook of Bilingualism: Psycholinguistic Approaches.* Oxford: Oxford University Press, pp. 89–108.

Dell, G. (1986). A spreading activation theory of retrieval in sentence production. *Psychological Review, 93,* 283-321.

Dell, Gary S., & O'Sheaghda, P. (1992). Stages of lexical access in language production. *Cognition, 42,* 287-314.

Deutsch A., & Bentin, S. (2001). Syntactic and semantic factors in processing gender agreement in Hebrew: Evidence from ERPs and eye movements. *Journal of Memory and Language, 45*(2), 200-224.

Dohmes, P., Zwitserlood, P. & Bölte, J. (2004). The impact of semantic transparency of morphologically complex words on picture naming. *Brain and Language. 90,* 203 - 212.

Donchin, E. (1979). Event-related brain potentials: A tool in the study of human information processing. In H. Begleiter (Ed.), Evoked potentials and behaviour. New York: Plenum.

Donchin, E., Ritter, W., & McCallum, C. (1978). Cognitive psychophysiology: The endogenous components of the ERP. In E. Callaway, P. Tueting, & S. Koslow (Eds.), Brain event-related potentials in man (pp. 349-441). New York: Academic Press.

Duyck, W., Van Assche, E., Drieghe, D., & Hartsuiker, R. J. (2007). Visual word recognition by bilinguals in a sentence context: evidence for non-selective lexical access. *Journal of Experimental Psychology: Learning, Memory and Cognition, 33*, 663-679.

Ellis, N. (2002). Frequency effects in language processing (a review with implications for theories of implicit and explicit language acquisition). *Studies in Second Language Acquisition, 24*, 143-188.

Ellis, R. (1984). *Classroom second language development*. Oxford: Pergamon.

Elman, J. L., & McClelland, J. L. (1988). Cognitive penetration of the mechanisms of perception: Compensation for coarticulation of lexically restored phonemes. *Journal of Memory and Language, 27,* 143-165.

Elston-Güttler, K. E., Gunter, T. C. & Kotz, S. A. (2005). Zooming into L2: Global language context and adjustment affect processing of interlingual homographs in sentences. *Cognitive Brain Research, 25* (1), 57-70.

Faussart, C., Jakubowicz, C., & Costes, M. (1999). Gender and number processing in spoken French and Spanish. *Rivista di Linguistica, 11,* 75–101.

Felser, C. & L. Roberts (2007). Processing *wh*-dependencies in a second language: A cross-modal priming study. *Second Language Research 23*, 9-36.

Fiorentino, R. (2006). Masked morphological priming of compound constituents: Specifying the locus of morpho-orthographic segmentation effects using English compound words. Unpublished raw data.

Flege, J., Yeni-Komshian, G. & Liu, S. (1999). Age Constraints on Learning L2 Phonology and Morphosyntax. *Proceedings* of *the Joint Meeting* of *the 16th*

International Congress on Acoustics and the 137th Meeting of *the Acoustical Society* of *America,* Berlin, 15-19 March, 1999.

Fodor, J. A. (1983). Modularity of Mind: An Essay on Faculty Psychology. Cambridge, Mass.: MIT Press.

Frauenfelder, U. H., & Tyler, L. K. (1987). Spoken word recognition. Cambridge: MIT Press.

Friederici, A. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences, 6*(2), 78-84.

Friederici, A. D. and Jacobsen, T. (1999). Processing grammatical gender during language comprehension. *Journal of Psycholinguistic Research*, *28*(5), 467–484.

Frigo, L., & McDonald, J. L. (1998). Properties of phonological markers that affect the acquisition of gender-like subclasses. *Journal of Memory and Language, 39,* 218-245.

García-Albea, J. E., Sanchez-Casas, R. & Valero, T. (1996, September). Form and meaning contribution to word recognition in Catalan-Spanish bilinguals. Paper presented at the meeting of the Ninth Conference of the European Society for Cognitive Psychology, University of Würzburg, Germany.

Gilhooly, K. J. and Logie, R. H. (1980). Age of acquisition, imagery, concreteness, familiarity and ambiguity measures for 1944 words. *Behaviour Research Methods and Instrumentation, 12,* 395-427.

Glushko, R. J. (1979). The organization and activation of orthographic knowledge in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance, 5,* 674-691.

Gollan, T. H., Forster, K. I., & Frost, R. (1997). Translation priming with different scripts: Masked priming with cognates and noncognates in Hebrew-English bilinguals. *Journal of Experimental Psychology-Learning Memory and Cognition, 23*(5), 1122-1139.

Gómez R. & Gerken L. A. (2000). Infant artificial language learning and language acquisition. *Trends in Cognitive Sciences, 4*, 178-86.

Greenhouse S W & Geisser S. (1959). On methods in the analysis of profile data. *Psychometrika* 24:95-112.

Guillelmon, D. and Grosjean, F. (2001). The gender marking effect in spoken word recognition: The case of bilinguals. *Memory and Cognition, 29*(3), 503-511.

Gumnior, H., Bölte, J., & Zwitserlood, P. (2006). A chatterbox is a box: Morphology in German word production. Language and Cognitive Processes, 21, 920–944.

Gunter, T., Friederici, A., & Schriefers, H., (2000). Syntactic gender and semantic expectancy: ERPs reveal early autonomy and late interaction. *Journal of Cognitive Neuroscience*, *12,* 556-568.

Hagoort, P., & Brown, C. (1999). Gender electrified: ERP evidence on the syntactic nature of gender processing. *Journal of Psycholinguistic Research*, *28*(6):715–728.

Hagoort, P., Wassenaar, M. E. D., & Brown, C. M. (2003). Syntax-related ERP-effects in Dutch. *Cognitive Brain Research, 16* (1), 38-50.

Hahne, A., & Friederici, A.D. (2001). Processing a second language: Late learners' comprehension mechanisms as revealed by event-related brain potentials. *Bilingualism: Language and Cognition, 4*, 123-141.

Hahne, A., Mueller, J. & Clahsen, H. (2006). Morphological processing in a second language: Behavioural and event-related potential evidence for storage and decomposition. *Journal of Cognitive Neuroscience, 18*, 121-134.

Harley, B., & Wang, W. (1997). The critical period hypothesis: Where are we now? In A. M. B. de Groot & J. F. Kroll (Eds.), *Tutorials in bilingualism: Psycholinguistic perspectives* (pp. 19-51). Mahwah, NJ: Erlbaum.

Harris, J. W. (1991). The exponence of gender in Spanish. *Linguistic Inquiry, 22,* 27-62.

Hawkins, R., & Chan, C. (1997). The partial availability of Universal Grammar in second language acquisition: The 'failed functional features hypothesis.' *Second Language Research, 13*, 187-226.

Hawkins, R., & Liszka, S. (2003). Locating the Source of Defective Past Tense Marking in Advanced L2 English Speakers. In R. van Hout, H. Aafke, F. Kuiken and R. Towell (eds). *The Interface Between Syntax and Lexicon in Second Language Acquisition* (pp. 21-44). Amsterdam: John Benjamins.

Hawkins, R., & Franceschina, F. (2004). Explaining the acquisition and nonacquistion of determiner noun gender concord in French and Spanish. In Prévost, P. & Paradis, J. (eds.). *The Acquisition of French in Different Contexts: Focus on functional categories* (pp 175 – 205). Philadelphia: John Benjamins Publishing Company.

Heim, S. (2005). The structure and dynamics of normal language processing. *Acta Neurobiologiae Experimentalis* 65, 95-116.

Hermans, D., Bongaerts, T., De Bot, K., & Schreuder, R. (1998). Producing words in a foreign language: Can speakers prevent interference from their first language? *Bilingualism: Language and Cognition, 1*(3), 213-229.

Hernandez, A. E., Martinez, A., & Kohnert, K. (2000). In search of the language switch: An fMRI study of picture naming in Spanish-English bilinguals. *Brain and Language, 73,* 421-431.

Hernandez, A., Bates, E., & Avila L. X. (1996). Processing across the language boundary: A cross-modal priming study of Spanish-English bilinguals. Journal of Experimental Psychology: Learning, Memory, and Cognition, 22(4), 846-864. Memory, and Cognition, 22(4), 846-864.

Hillyard, S. A., & Picton, T. W. (1987). Electrophysiology of cognition. In . F. Plum (ed.), *Handbook of Physiology: Section 1. The Nervous System: Vol. 5. Higher Functions of the Brain, Part 2,* (pp. 519-584). Bethesda, Md.: Waverly Press.

Hofmann, J. (2005). Phonological, Morphological, and Semantic Aspects of Grammatical Gender Processing in German. Unpublished Ph.D. dissertation, Max Planck Institute in Human Cognitive and Brain Sciences, Leipzig, Germany.

Hohlfeld, A. (2006). Accessing grammatical gender in German: The impact of gender-marking regularities. Applied Psycholinguistics, 27, 127-142.

Holcomb, P. J. (1993). Semantic priming and stimulus degradation: implications for the role of the N400 in language processing. *Psychophysiology, 30,* 47–61.

Holcomb, P. J., & Grainger, J. (2006). On the time course of visual word recognition: An event-related potential investigation using masked repetition priming. Journal of Cognitive Neuroscience, 18, 1631-1643.

Hoshino, N., Dussias, P. E., & Kroll, J. F. (under revision). The role of conceptual resources in guiding the production of subject-verb agreement in bilingual and monolingual speakers.

Hyltenstam, K. & Abrahamsson, N. (2001). Age and L2 learning: The hazards of matching practical "implications" with theoretical "facts". (Comments on Stefka H. Marinova-Todd, D. Bradford Marshall, and Catherine E. Snow's "Three misconceptions about age and L2 learning"). *TESOL Quarterly, 35:1*, 151-170.

Isel, F., Gunter, T., & Friederici, A. (2003). Prosody-assisted head-driven access to spoken German compounds. Journal of Experimental Psychology: Learning, Memory, and Cognition, 29, 277–288.

Jackendoff, R. (2002). *Foundations of language: Brain, meaning, grammar, evolution*. Oxford, UK: Oxford University Press.

Jared, D. & Kroll, J. F. (2001). Do bilinguals activate phonological representations in one or both of their languages when naming words? Journal of Memory and Language, 44, 2-31.

Jiang, N. (2000). Lexical representation and development in a second language. *Applied Linguistics, 21,* 47-77.

Jiang, N. (2004). Semantic Transfer and Its Implications for Vocabulary Teaching in a Second Language. *The Modern Language Journal, 88 (3) 416-432.*

Jiang, N. (2007). Selective integration of linguistic knowledge in adult second language learning. *Language Learning,* 57, 1-33.

Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology, 21,* 60–99.

Juhasz, B. J., Starr, M. S., Inhoff, A. & Placke, L. (2003). The effects of morphology on the processing of compound words: Evidence from lexical decision, naming and eye fixations. *British Journal of Psychology*, 94, 223-244.

Karmiloff-Smith, A. (1978). The interplay between syntax, semantics and phonology in language acquisition processes. In R. N.Campbell & P. T.Smith (Eds.), *Advances in the Psychology of Language*, (pp. 1-23). London: Plenum Press.

Kim, K., Relkin, N., Lee, K., and Hirsch, J. (1997). Distinct cortical areas associated with native and second languages. *Nature*, 388, 171--174.

Klein, D., Milner, B., Zatorre, R., Meyer, E., & Evans, A. (1995). The neural substrates underlying word generation: A bilingual functional-imaging study. *Proceedings of the National Academy of Sciences U.S.A., 92*, 2899-2903.

Klein, K., & Fiss, W. H. (1999). The reliability and stability of the Turner and Engle working memory task. *Behavior Research Methods, Instruments and Computers, 31*, 429–432.

Kluender, R., & Kutas, M. (1993). Bridging the gap: Evidence from ERPs on the processing of unbounded dependencies. *Journal of Cognitive Neuroscience, 5*(2), 196-214.

Knightly, L. M., Jun, S.-A., Oh, J. S., & Au, T. K. (2003). Production benefits of childhood overhearing. *Journal of the Acoustical Society of America, 114*(1), 465-474.

Koester D., Gunter, Th. C., Wagner, S., & Friederici, A. D. (2004). Morphosyntax, prosody, and linking elements: the auditory processing of German nominal compounds. *Journal of Cognitive Neuroscience, 16*(9), 1647–1668.

Kotz, S., Holcomb, P. J., & Osterhout, L. (2008). ERPs reveal comparable syntactic sentence processing in early bilinguals and monolinguals. Acta Psychologica, 128(3), 514-27.

Kotz, S. A., & Elston-Güttler, K. E. (2004). The role of proficiency on processing categorical and associative information in the L2: Reaction times and event-related potentials. *Journal of Neurolinguistics, 17*, 215-235.

Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. Journal of Memory and Language, 33, 149-174.

Kučera, F. & Francis, W. (1967). *Computational analysis of present-day American English.* Providence, RI: Brown University Press.

Kuchenbrandt, I. (2008). Cross-linguistic influences in the acquisition of grammatical gender? *Working Papers in Multilingualism, 86,* Series B.

Kupisch, T. (2007). Determiners in bilingual German-Italian children: What they tell us about the relation between language influence and language dominance. *Bilingualism: Language and Cognition 10*(1). 57-78.

Kutas, M. & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Science, 4*, 463-470.

La Pointe, L. B., & Engle, R. W. (1990). Simple and complex word spans as measures of working memory capacity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 1118–1133.

LaHeij, W., De Bruyn, E., Elens, E., Hartsuiker, R., Helaha, D., & vanSchelven, L. (1990). Orthographic facilitation and and categorical interference in a word-translation variant of the Stroop task. *Canadian Journal of Psychology, 44*, 76-83.

Lehtonen, M. & Laine, M. (2003). How word frequency affects morphological processing in monolinguals and bilinguals. *Bilingualism: Language and Cognition*, 6, 213-225.

Lehtonen, M., Vorobyev, V., Soveri, A., Hugdahl, K., Tuokkola, T., & Laine, M. (2007). Neural correlates of morphological decomposition in a morphologically rich language: An fMRI study. *Brain and Language 98*(2): 182-193.

Lenneberg, E. H. (1967). *Biological foundations of language*. New York: Wiley.

Levelt, W. (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), 1-38.

Levelt, W. J. M., Schriefers, H., Vorberg, D., Meyer, A. S., Pechmann, T., & Havinga, J. (1991). The time course of lexical access in speech production: A study of picture naming. *Psychological Review, 98*, 122-142.

Lew-Williams, C., & Fernald, A. (2007). Young children learning Spanish make rapid use of grammatical gender in spoken word recognition. *Psychological Science, 33*, 193-198.

Libben, G. (2006). Why study compound nouns? An overview of the issues. In *The Representation and Processing of Compound Nouns*. Eds. G. Libben & G. Jarema. Oxford University Press.

Libben, G., Derwing, B., & de Almeida, R. G. (1999) Ambiguous Novel Compounds and Models of Morphological Parsing. *Brain and Language, 68*, 378-386.

Liceras, J. M. 1997. El saber sí ocupa lugar: la ciencia cognitiva y la adquisición del español como lengua segunda. Revista Complutense de Educación 14: 299-318.

Linck, J. A., Sunderman, G., & Kroll, J. F. (in preparation). Losing access to the native language while immersed in a second language: Evidence for the role of inhibition in second language learning.

Lleó, C. & K. Demuth (1999). Prosodic Constraints on the Emergence of Grammatical Morphemes: Crosslinguistic Evidence from Germanic and Romance Languages. In A. Greenhill, H. Littlefield & Ch. Tano (eds.), *Proceedings of the 23rd Annual Boston University Conference on Language Development, Vol. 2*, 407-418. Somerville, Ma: Cascadilla Press.

Long, M. H. (1990). Maturational constraints on language development. *Studies in Second Language Acquisition, 12*(3), 251-85.

Lozano, C. 2003. Universal Grammar and focus constraints: The acquisition of pronouns and word order in non-native Spanish. University of Essex: Unpublished PhD dissertation.

Lu, C.-H., & Proctor, R. W. (1995). The influence of irrelevant location information on performance: A review of the Simon and spatial Stroop effects. *Psychological Bulletin and Review, 2*, 174-207.

Luck, S. J. (2005). An introduction to the event-related potential technique. Cambridge, MA: MIT Press.

MacDonald, M. C., Pearlmutter, N. J. & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review, 101*, 676-703.

MacWhinney, B. (1987). The Competition Model. In B. MacWhinney (Ed.), *Mechanisms of language acquisition*. Hillsdale, NJ: Lawrence Erlbaum.

MacWhinney, B. (1997). Second language acquisition and the Competition Model. In J. Kroll & A. De Groot (Eds.), *Tutorials in bilingualism*, Mahwah, NJ: Lawrence Erlbaum.

MacWhinney, B. (2006). Emergentism -- use often and with care. *Applied Linguistics, 27*(4), 729-740.

MacWhinney, B. (1989). 'Competition and lexical categorization' in R. Corrigan, F. Eckman, and M. Noonan (eds): Linguistic Categorization (pp. 195–242). Philadelphia: Benjamins.

Marcus, G. F. (1999). Connectionism: with or without rules? Response to J. L. McClelland and D. C. Plaut (1999), *Trends in Cognitive* Sciences, 3, 168-170.

Marcus, G. F. (2001). The Algebraic Mind: Integrating Connectionism and Cognitive Science. Cambridge, MA: MIT Press.

Marcus, G. F., Brinkmann, U., Clahsen, H., Wiese, R. & Pinker, S. (1995). German inflection: the exception that proves the rule. *Cognitive Psychology 29*: 189-256.

Marinova-Todd, S. H., Marshall, D. B., & Snow, C. E. (2000). Three misconceptions about age and second language acquisition. *TESOL Quarterly, 34, 9-34*.

Marinova-Todd, S. H., Marshall, D. B., & Snow, C. E. (2001). Missing the point: A response to Hyltenstam and Abrahamsson. *TESOL Quarterly, 35, 171-76*.

Marslen-Wilson, W. D., Tyler, L. K., Wakaler, R., & Older, L. (1994). Morphology and meaning in the English mental lexicon. *Psychological Review, 101,* 3-33.

Maye, J., Werker, J. F., & Gerken, L. A. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, *82*(3), B101–B111.

McClelland, J. L., & Plaut, D. C. (1999). Does generalization in infant learning implicate abstract algebra-like rules? Trends in Cognitive Sciences, 3, 166-168.

McDonald, J. L. (2006). Beyond the critical period: Processing-based explanations for poor grammaticality judgment performance by late second language learners. *Journal of Memory and Language*, *55,* 381-401.

McLaughlin, J., Osterhout, L., & Kim, A. (2004). Neural correlates of second-language word learning: minimal instruction produces rapid change. Nature Neuroscience, 7, 703-704.

Michael, E., & Gollan, T. H. (2005). Being and becoming bilingual: Individual Differences and consequences for language production. In J. F. Kroll & A. M. B. de Groot (Eds.), *The* h*andbook of bilingualism: Psycholinguistic approaches* (pp. 389-407). New York: Oxford University Press.

Mills, A. E. (1986). The acquisition of gender: a study of German and English. *Springer Series in Language and Communication; 20*. Berlin: Springer.

Minagawa-Kawai, Y., Mori, K. and Sato, Y. (2005). Different brain strategies underlie the categorical perception of foreign and native phonemes. *Journal of Cognitive Neuroscience 17* (9): 1376-1385.

Misra, M., Guo, T., Bobb, S. C., & Kroll, J. F. (in preparation). ERP Evidence for Inhibition in Bilingual Word Production.

Mitchell, D. C., Cuetos, F., Corley, M. M. B. & Brysbaert, M. (1995). Exposure-based models of human parsing: Evidence for the use of coarse-grained (non-lexical) statistical records. *Journal of Psycholinguistic Research, 24*(6), 469-488.

Möhring, Anja. 2001. The acquisition of French by German pre-school children: An empirical investigation of gender assignment and gender agreement. EUROSLA Yearbook, Foster-Cohen, Susan H. and Anna Nizegorodcew (eds.), 171–193.

Molfese, D., Molfese, V., & Kelly, S. D. (2001). The use of brain electrophysiology techniques to study language and reading: A basic guide for the beginning consumer of electrophysiology information. Learning Disabilities Quarterly, 24(3), 177-188.

Montrul, S. & Slabakova, R. (2003). Competence similarities between native and near-native speakers: An investigation of the Preterite/Imperfect contrast in Spanish. *Studies in Second Language Acquisition, 25*(3): 351-398.

Newport, E. L., & Aslin, R. N. (2004). Learning at a distance: I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology, 48*, 127-162.

Oh, J. S., Jun, S.-A., Knightly, L. M., & Au, T. K. (2003). Holding on to childhood language memory. *Cognition, 86*(3), B53-B64.

Oliphant, K. (1998). Acquisition of grammatical gender in Italian as a foreign language. *The Canadian Modern Language Review/La revue canadienne des langues vivantes*, *54*(2), 237-262.

Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*, 31, 785-806.

Osterhout, L., McLaughlin, J., Inoue, K., & Loveless, J. (2000). *Brain responses in L2 learning.* Paper presented at the meeting: Bilingualism: From basic research to educational practice, University of Trento, Rovereto, Italy.

Pallier, C., Dehaene, S., Poline, J.-B., LeBihan, D., Argenti, A.-M., Dupoux, E., & Mehler, J. (2003). Brain imaging of language plasticity in adopted adults: can a second language replace the first? *Cerebral Cortex*, 13(2):155-161.

Pallier, C., Bosch, L., & Sebastian-Gallés, N. (1997). A limit on behavioral plasticity in speech perception. Cognition, 64(3):B9-B17.

Penfield, W., & Roberts, L. (1959). *Speech and brain-mechanisms*. Princeton, NJ: Princeton University Press.

Perani, D., Abutalebi, J., Paulesu, E., Brambati, S., Scifo, P., Cappa, S. F., et al. (2003). The role of age of acquisition and language usage in early, high-proficient bilinguals: An fMRI study during verbal fluency. *Human Brain Mapping*, *19*, 170–182

Perani, D., Dehaene, S., Grassi, F., Cohen, L., Cappa, S. F., Dupoux, E., et al. (1996). Brain processing of native and foreign languages. *Neuroreport*, *7*(15–17), 2439–2444.

Perani, D., Paulesu, E., Galles, N. S., Dupoux, E., Dehaene, S., Bettinardi, V., et al. (1998). The bilingual brain. Proficiency and age of acquisition of the second language. *Brain*, *121*, 1841–1852.

Pierrehumbert, J. (2003). Phonetic diversity, statistical learning, and acquisition of phonology, Language and Speech, 46(2-3), 115-154.

Pinker, S. & Ullman, M. (2002) The past and future of the past tense. *Trends in Cognitive Science, 6,* 456-463.

Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology, 118*, 2128-2148.

Portin, M., Lehtonen, M., & Laine, M. (2006). Processing of inflected nouns in late bilinguals. *Applied Psycholinguistics 28*, 135-156.

Posner, M. I. & Snyder, C. R. R. (1975). Attention and cognitive control. In R. L. Solso (Ed.), *Information processing and cognition: The Loyola symposium* (pp. 5585). Hillsdale, NJ: Erlbaum.

Quasthoff, U. (2002). Deutscher Wortschatz im Internet [Online database]. http://www.wortschatz.uni-leizpig.de/ (2002). Leipzig, Germany: University of Leipzig.

Rayner, K. & Pollatsek, A. (1987). Eye movements in reading: A tutorial review. In M. Coltheart (Ed.), *Attention & Performance: Vol. 12. The psychology of reading* (pp. 327-362). Hillsdale, NJ: Lawrence Erlbaum Associates.

Regan, D. (1989). *Human brain electrophysiology. Evoked potentials and evoked magnetic fields in science and medicine*. New York: Elsevier.

Roelofs, A., & Baayen, H. (2002). Morphology by itself in planning the production of spoken words. *Psychonomic Bulletin & Review, 9*, 132-138.

Rossi, S., Gugler, M.F., Friederici, A. D., & Hahne, A. (2006). The impact of proficiency on syntactic second-language processing of German and Italian: evidence from event-related potentials. *Journal of Cognitive Neuroscience, 18,* 2030-2048.

Rueschemeyer, S.-A., Fiebach, C. J., Kempe, V., & Friederici, A. D. (2005). Processing lexical-semantic and syntactic information in first and second language: fMRI evidence from German and Russian. *Human Brain Mapping, 25,* 266- 286.

Rumelhart, D. E., & McClelland, J. L. (1986). On learning the past tense of English verbs. In J. L. McClelland, D. E. Rumelhart and PDP Research Group (Eds), *Parallel distributed processing: Volume 2* (pp. 216-271). Cambridge, MA: MIT Press.

Sabourin, L. (2006). Does the "shallow-structures" proposal account for qualitative differences in L1 and L2 processing. *Journal of Applied Psycholinguistics*. *27*(1), 81-84.

Sabourin, L., Stowe, L., & de Haan, G. J. (2006). Transfer effects in learning a second language grammatical gender system. *Second Language Research, 22,* 1-29.

Saffran, J. R. (2003). Statistical language learning: Mechanisms and constraints. *Current Directions in Psychological Science, 12*, 110-114.

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition, 70,* 27-52.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month old infants. Science, 274, 1926-1928.

Sánchez-Casas, R., & García-Albea, J. E. (2005). The representation of cognate and noncognate wordson bilingual memory: Can cognate status be characterized as a special kind of morphological ralation? J. F. Kroll, A.M.B. de Groot, *Handbook of Bilingualism*, (pp. 226-250).

Scherag, A., Demuth, L., Rösler, F., Neville, H., & Röder, B. (2004). The effects of late acquisition of L2 and the consequences of immigration on L1 for semantic and morpho-syntactic language aspects. *Cognition, 93,* B97-B108.

Schiller, N. O., & Caramazza, A. (2003). Grammatical feature selection in noun phrase production: Evidence from German and Dutch. *Journal of Memory and Language, 48,* 169-194.

Schiller, N. O., Münte, T., Horemans, I., & Jansma, B. M. (2003). The influence of semantic and phonological factors on syntactic decisions: An event-related brain potential study. *Psychophysiology,40,* 869-877.

Schneider, W., Eschmann, A., & Zuccolotto, A. (2002). E-Prime v1.1. Pittsburgh, PA: Psychology Software Tools Inc.

Schreuder, R., & Baayan, R. H. (1995). Modeling morphological processing. In L. Feldman (Ed.), Morphological aspects of language processing (pp. 131-154). Hillsdale, NJ: Lawrence Erlbaum.

Schriefers, H., Friederici, A., & Rose, (1998). Context effects in visual word recognition: Lexical relatedness and syntactic context. *Memory & Cognition, 26,* 1292-1303.

Schwartz, B. D., & Sprouse, R. (1994). Word order and nominative case in nonnative language acquisition. In T. Hoekstra & B. Schwartz (Eds), *Language acquisition studies in generative grammar* (pp. 317 – 368). Amsterdam: Benjamins.

Schwartz, B. D., & Sprouse, R. (1996). L2 cognitive states and the full transfer/ full access model. *Second Language Research, 12,* 40-72.

Schwichtenberg, B. & Schiller, N. O. (2004). Semantic gender assignment regularities in German. *Brain and Language, 90,* 326-337.

Sebastián-Gallés, N. and Soto-Faraco, S. (1999). On-line processing of native and non-native phonemic contrasts in early bilinguals. *Cognition, 72*, 111-123.

Sebastián-Gallés, N., Echeverría, S., & Bosch, L. (2005). The influence of initial exposure on lexical representation: Comparing early and simultaneous bilinguals. *Journal of Memory and Language*, 52, 240-255.

Seidenberg, M. S. & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review, 96*, 523-568.

Sekerina, I. A., & Pugach, Y. (2005). Cross-linguistic variation in gender use as a parsing constraint: Dutch vs. Russian. In Franks, S., Gladney, F. Y., & Tasseva-Kurktchieva, M. (Eds.), The Proceedings of the 13th Annual Workshop on Formal Approaches to Slavic Linguistics. The South Carolina Meeting 2004. 312-323. Michigan Slavic Publications.

Shoolman, N. & Andrews, S. (2003). Racehorses, reindeers and sparrows: Using masked priming to investigate morphological influences on word identification. In S. Kinoshita & S. Lupker (Eds.*), Masked priming: The state of the art*. Psychology Press.

Silva, R. & Clahsen, H. 2008. Morphologically complex words in L1 and L2 processing: Evidence from masked priming experiments in English. *Bilingualism: Language and Cognition 11*, 245-260.

Silverberg, S., & Samuel, A. G. (2004). The effect of age of second language acquisition on the representation and processing of second language words. *Journal of Memory and Language, 51*, 381-398.

Simon, J. R. and Rudell, A. P. (1967). Auditory S-R compatibility: The effect of an irrelevant cue on information processing. *Journal of Applied Psychology, 51*, 300-304.

Slobin, D., & Bever, T. (1982). A cross-linguistic study of sentence comprehension. *Cognition, 12,* 229-265.

Sorace, A. (2003). Near-nativeness. In M. Long and C. Doughty (eds.), *Handbook of Second Language Acquisition*, 130-152. Oxford: Blackwell.

Squires, K. C., Squires, N. K., & Hillyard, S. A. (1975). Decision related cortical potentials during an auditory signal detection task with cued observation intervals. *Journal of Experimental Psychology: Human Perception & Performance, 104,* 268-279.

Stemberger, J.-P. (2001). Morphology in language production with special reference to connectionism. In Spencer, A. & Zwicky, A. *The Handbook of Morphology*. Oxford: Blackwell Publishers.

Sunderman, G., & Kroll, J. F. (2006). First language activation during second language lexical processing: An investigation of lexical form, meaning, and grammatical class. *Studies in Second Language Acquision.*

Swinney, D. (1979). Lexical access during sentence comprehension: (Re) consideration of context effects. *Journal of Verbal Learning and Verbal Behavior, 18*, 645-659.

Taft, M. (2004). Morphological decomposition and the reverse base frequency effect. *The Quarterly Journal of Experimental Psychology, 57A (4)*, 745-765.

Taft, M., & Forster, K. I. (1976). Lexical storage and retrieval of polymorphemic and polysyllabic words. Journal of Verbal Learning and Verbal Behavior, 15, 607-620.

Talamas, A., Kroll, J. F., & Dufour, R. (1999). From form to meaning: Stages in the acquisition of second-language vocabulary. *Bilingualism: Language and Cognition*, *2,* 45–58.

Tanenhaus, M. K., Dell, G. S. & Carlson, G. (1987). Context effects in lexical processing: A connectionist approach to modularity. In: J. Garfield (Ed.), *Modularity in knowledge representation and natural language understanding*, (pp. 83-108). Cambridge, MA: MIT Press.

Taraban, R., & Kempe, V. (1999). Gender processing in native and non-native Russian speakers. *Applied Psycholinguistics, 20,* 119-148.

Tokowicz, N., & Kroll, J. F. (2007). Number of meanings and concreteness: Consequences of ambiguity within and across languages. Language and Cognitive Processes, 22, 727-779.

Tokowicz, N., & MacWhinney, B. (2005). Implicit and explicit measures of sensitivity to violations in second language grammar. *Studies in Second Language Acquisition, 27*, 173-204.

Tokowicz, N., Michael, E., & Kroll, J. F. (2004). The roles of study abroad experience and working memory capacity in the types of errors made during translation. *Bilingualism: Language and Cognition, 7,* 255-272.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent?. *Journal of Memory and Language, 28,* 127-154.

Ullman, M. T. (2004) Contributions of memory circuits to language: the declarative/procedural model. *Cognition*, *92*, 231–270.

van Berkum, J. J., Brown, C. M., and Hagoort, P. (1999). When does gender constrain parsing? Evidence from ERPs. *Journal of Psycholinguistic Research*, *28*(5), 555–571.

van Hell, J. G. (1998). *Cross-language processing and bilingual memory organization*. Unpublished doctoral dissertation, University of Amsterdam.

van Hell, J. G., & de Groot, A. M. B. (1998). Disentangling context availability and concreteness in lexical decision and word translation. *Quarterly Journal of Experimental Psychology, 51A*, 41-63.

Van Petten, C., & Luka, B. J. (2006). Neural bases of semantic context effects in electromagnetic and hemodynamic studies. *Brain and Language. 97*, 279-293.

Ventureyra, V., Pallier, C. & Yoo, H. Y. (2004). “The loss of first language phonetic perception in adopted Koreans”. *Journal of Neurolinguistics*, 17, 7991.

Vigil-Colet, A., Pérez-Ollé, J. & García-Albea, J. E. (2000). The role of the P300 component in a translation-recognition task. *Psicothema*, *12* (4), 605-614.

Vosse, T. & Kempen, G. (2000). Syntactic structure assembly in human parsing: A computational model based on competitive inhibition and a lexicalist grammar. *Cognition*, *75*, 105-143.

Wartenburger, I., Heekeren, H. R., Abutalebi, J., Cappa, S. F., Villringer, A., & Perani, D. (2003). Early setting of grammatical processing in the bilingual brain. *Neuron*, *37*, 159-170.

Weber-Fox, C. M. and Neville, H. J. (1996). Maturational constraints on functional specializations for language processing: ERP and behavioral evidence in bilingual speakers. *Journal of Cognitive Neuroscience 8*(3):231-256.

White, L., & Genesee, F. (1996). How native is near-native? The issue of ultimate attainment in adult second language acquisition. *Second Language Research, 17,* 233–265.

Wilkins, W., & Wakefield, J. (1995). Brain evolution and neurolinguistic preconditions. Behavioral and Brain Sciences, 18, 161-182.

Zwitserlood, P. (1994). Processing and representation of Dutch compounds: Effects of semantic transparency. *Language and Cognitive Processes, 9* (3), 341 - 368.

Zwitserlood, P., Bölte, J., & Dohmes, P. (2000). Morphological effects on speech production; Evidence from picture naming. Language and Cognitive Processes, 15, 563–591.

Appendix A

Critical Items for Picture Naming Task

Picture	Item Purpose	English name	Length	Log Freq[a]	German name	Length	Log Freq[a]
AIRPLANE	experimental	airplane	8	1.9	flugzeug	8	4.3
APPLE	practice	apple	5	3.4	apfel	5	2.6
BAT	practice	bat	3	2.7	fledermaus	10	1.4
BELT	experimental	belt	4	3.3	guertel	6	2.1
BICYCLE	experimental	bicycle	7	1.8	fahrrad	7	4.9
BIRD	experimental	bird	4	0.0	vogel	5	3.7
BOOK	practice	book	4	6.1	buch	4	5.2
BOTTLE	experimental	bottle	6	4.8	flasche	7	3.6
BOY	experimental	boy	3	5.9	junge	5	4.2
BREAD	experimental	bread	5	4.3	brot	4	3.4
BUTTERFLY	experimental	butterfly	9	2.4	schmetterling	13	1.4
BUTTON	experimental	button	6	3.3	knopf	5	2.5
CAKE	experimental	cake	4	3.6	torte	5	1.1
CAR	experimental	car	3	5.9	auto	4	4.8
CHAIR	experimental	chair	5	4.9	stuhl	5	3.7
CHEESE	practice	cheese	6	3.5	kaese	4	2.1
CHERRY	experimental	cherry	6	2.1	kirsche	7	1.4
CHURCH	experimental	church	6	5.2	kirche	6	5.4
CLOCK	experimental	clock	5	3.7	uhr	3	6.6
COW	experimental	cow	3	3.7	kuh	3	3.7
CROWN	practice	crown	5	3.2	krone	5	3.3
CUP	experimental	cup	3	4.4	tasse	5	2.4
DOG	experimental	dog	3	4.8	hund	4	4.1
DOOR	experimental	door	4	6.0	tuer	3	4.8
DRESS	experimental	dress	5	4.5	kleid	5	3.7
EAR	experimental	ear	3	4.5	ohr	3	3.9
EGG	experimental	egg	3	4.5	ei	2	3.9
EYE	experimental	eye	3	6.3	auge	4	5.7
FLOWER	experimental	flower	6	4.5	blume	5	3.5
FOOT	experimental	foot	4	5.8	fuss	3	4.6
FORK	experimental	fork	4	2.8	gabel	5	1.6
GIRL	experimental	girl	4	6.1	maedchen	7	5.4
GLASSES	experimental	glasses	7	3.5	brille	6	2.9
GLOVE	experimental	glove	5	3.0	handschuh	9	2.3
GRAPES	experimental	grapes	6	0.0	weintrauben	11	0.0
HAIR	practice	hair	4	5.3	haare	5	4.3
HAT	practice	hat	3	4.2	hut	3	3.0
HEART	experimental	heart	5	5.1	herz	4	5.0
HORSE	experimental	horse	5	4.9	pferd	5	4.2

KEY	experimental	key	3	4.5	schluessel	9	3.3
KNIFE	experimental	knife	5	3.8	messer	6	2.2
LEAF	experimental	leaf	4	4.4	blatt	5	4.3
LEG	experimental	leg	3	5.2	bein	4	4.3
LEMON	experimental	lemon	5	2.8	zitrone	7	1.4
MIRROR	experimental	mirror	6	3.9	spiegel	7	3.7
MONKEY	experimental	monkey	6	2.9	affe	4	2.1
MOUNTAIN	experimental	mountain	8	4.4	berg	4	4.2
ONION	experimental	onion	5	2.8	zwiebel	7	2.6
PANTS	experimental	pants	5	2.8	hose	4	3.2
PEANUT	experimental	peanut	6	1.8	erdnuss	6	0.0
PEAR	practice	pear	4	1.9	birne	5	1.6
PEN	experimental	pen	3	3.3	kugelschreiber	14	1.4
PENCIL	experimental	pencil	6	3.0	bleistift	9	1.8
PIANO	experimental	piano	5	3.3	klavier	7	2.8
PIG	experimental	pig	3	3.8	schwein	7	3.3
PLATE	experimental	plate	5	4.0	teller	6	2.5
POOL	experimental	pool	4	3.7	schwimmbecken	13	0.0
PURSE	experimental	purse	5	2.4	tasche	6	3.6
QUEEN	experimental	queen	5	4.0	koenigin	7	3.1
RABBIT	experimental	rabbit	6	3.0	hase	4	2.6
SHEEP	experimental	sheep	5	3.7	schaf	5	2.7
SHIRT	experimental	shirt	5	4.1	hemd	4	3.3
SKIRT	experimental	skirt	5	3.4	rock	4	2.4
SNAKE	experimental	snake	5	3.2	schlange	8	2.9
SPIDER	experimental	spider	6	2.1	spinne	6	1.9
SPOON	experimental	spoon	5	2.8	loeffel	6	1.9
STOVE	practice	stove	5	3.0	herd	4	2.9
STRAWBERRY	experimental	strawberry	10	1.9	erdbeere	8	1.1
SUN	experimental	sun	3	5.0	sonne	5	4.5
TABLE	experimental	table	5	5.5	tisch	5	4.6
TIE	practice	tie	3	3.6	schlips	7	1.4
TOOTHBRUSH	experimental	toothbrush	10	1.1	zahnbuerste	10	0.7
TRAIN	experimental	train	5	4.4	zug	3	4.8
TREE	experimental	tree	4	5.3	baum	4	4.2
UMBRELLA	experimental	umbrella	8	2.7	regenschirm	11	1.6
WINDOW	experimental	window	6	5.3	fenster	7	4.5

Notes .[a]Baayen, Pepenbrock, & Gulikers, 1995)

Appendix B

Language History Questionnaire

Subject Number: ______________________	Date: ______________________

This questionnaire is designed to give us a better understanding of your experience with other languages. We ask that you be as accurate and as thorough as possible when answering the following questions.

Part I

1. Gender: ___________________

2. Age: ______ years

3. Do you have any known visual and/or hearing problems (either corrected or uncorrected)?
   - ❑ No
   - ❑ Yes [Please explain: ____]

4. Native Country/Countries (Please check all that apply.)
   - ❑ United States
   - ❑ Other [Please specify: ___________________]

5. Native Language(s) (Please check all that apply.)
   - ❑ English
   - ❑ Other [Please specify: ___________________]

6. Language(s) spoken at home. (Please check all that apply.)
   - ❑ English
   - ❑ Spanish

- ❑ German
- ❑ Chinese
- ❑ Other [Please specify: ____]

Part II

The next section of the questionnaire deals with your second language learning experience.

7. Have you studied any second language(s)?
   - ❑ No → **If NO, please go to Part IV (on the final page of this questionnaire).**
   - ❑ Yes → If yes, which language(s)?

   ____

8. If you studied any second language(s) before college, please check all of the following that apply and indicate the starting age and length of study for any second language(s) learned before college.
   - ❑ Home/Outside of School – Language(s): ___________________________
     Starting age? __________ For how long? _____________
   - ❑ Elementary School – Language(s): ___________________________
     Starting age? __________ For how long? _____________
   - ❑ Middle School – Language(s): ___________________________
     Starting age? __________ For how long? _____________
   - ❑ High School – Language(s): ___________________________
     Starting age? __________ For how long? _____________

9. Have you studied any second language(s) in college?
   - ❑ No → **If NO, please go to Question # 13.**
   - ❑ Yes → If yes, which language(s)?

   ____

   For how long?

   - ❑ Less than one semester
   - ❑ 1-2 semesters
   - ❑ 3-4 semesters
   - ❑ 5-6 semesters
   - ❑ 7-8 semesters
   - ❑ 8+ semesters

10. Please list the most advanced second language course(s) you have completed in college:
____

11. Are you currently taking at least one second language course in college?
    - ❑ No
    - ❑ Yes → If yes, which course(s)? ____

12. Are you: (Please check all that apply and indicate which language each applies to if you have studied more than one second language at college.)
    - ❑ A Spanish, German, etc. 3 student.
    - ❑ Taking a second language for a requirement but interested in being a major or minor.
    - ❑ Taking a second language for a requirement; NOT interested in being a major or minor.
    - ❑ A second language minor.
    - ❑ A second language major.
    - ❑ A second language graduate student.
    - ❑ Other [Please explain: ____]

13. Have you studied and/or lived abroad?
    - ❑ Yes
    - ❑ No

If YES, where and when did you study, for how long, and what language(s) did you speak?

Country	Approx. dates	Length of Stay	Language

14. What do you consider to be your primary second language? (You may check more than one if you feel that you have multiple "primary" second languages.)
    - ❑ English
    - ❑ Spanish
    - ❑ German
    - ❑ Chinese
    - ❑ Other [Please specify: ____]

15. What language do you currently think is your *dominant* language (i.e., the language you are most comfortable using on a daily basis)? (Please check one)
    - ❑ English
    - ❑ Spanish
    - ❑ German
    - ❑ Chinese
    - ❑ Other [Please specify: ____]

Part III

The next section asks you to rate your skills in your primary second language. If English is your primary second language, then rate yourself on your native language in this section (you will rate your English skills later). If you have more than one "primary" second language, please indicate your skills for each language separately by writing the language next to the number that matches your skill level.

What language(s) are these ratings for? ______________________________________

16. Your reading proficiency in this language. (1=not literate and 10=very literate)

1 2 3 4 5 6 7 8 9 10

17. Your spelling proficiency in this language. (1=not good and 10=very good)

1 2 3 4 5 6 7 8 9 10

18. Your writing proficiency in this language. (1=not literate and 10=very literate)

1 2 3 4 5 6 7 8 9 10

19. Your speaking ability in this language. (1=not fluent and 10=very fluent)

1 2 3 4 5 6 7 8 9 10

20. Your speech comprehension ability in this language. (1=unable to understand conversation and 10=perfectly able to understand)

1 2 3 4 5 6 7 8 9 10

21. In my second language classes, I get:

- ❑ Mostly As
- ❑ Mostly As and Bs
- ❑ Mostly Bs
- ❑ Mostly Bs and Cs
- ❑ Mostly Cs

Part IV

The next section of the questionnaire deals with your English language skills. Please rate yourself on each measure by circling the appropriate number.

These ratings are for ENGLISH.

22. Your English reading proficiency. (1=not literate and 10 = very literate)

1 2 3 4 5 6 7 8 9 10

23. Your English spelling proficiency. (1=not good and 10=very good)

1 2 3 4 5 6 7 8 9 10

24. Your English writing proficiency. (1=not literate and 10=very literate)

1 2 3 4 5 6 7 8 9 10

25. Your English speaking ability. (1=not fluent and 10=very fluent)

1 2 3 4 5 6 7 8 9 10

26. Your English speech comprehension ability. (1=unable to understand conversation and 10=perfectly able to understand)

1	2	3	4	5	6	7	8	9	10

27. Do you have any additional comments to make? Please include any additional language experience that you have not included in other portions of this questionnaire.

Thank you for your participation!

Appendix C

Words and Equation Stimuli Used in the Operations Span Task

Trial Type	Set Size	Equation	English Word	German Word	Correct Response
practice	4	(8 / 4) + 3 = 1	father	Vater	no
practice	4	(3 * 4) - 3 = 9	uncle	Onkel	yes
practice	4	(16 / 4) + 2 = 2	money	Geld	no
practice	4	(12 / 3) - 2 = 2	degree	Diplom	yes
practice	6	(2 * 6) - 3 = 5	dollar	Dollar	no
practice	6	(2 * 5) - 2 = 8	wood	Holz	yes
practice	6	(4 * 3) - 4 = 4	metal	Metall	no
practice	6	(14 / 2) + 2 = 9	dog	Hund	yes
practice	6	(10 / 2) + 3 = 2	chest	Kasten	no
practice	6	(9 * 1) - 2 = 7	blood	Blut	yes
critical	2	(18 / 3) - 4 = 2	hotel	Hotel	yes
critical	2	(4 * 1) + 2 = 2	author	Autor	no
critical	2	(16 * 1) - 9 = 7	poem	Gedicht	yes
critical	2	(10 / 1) - 2 = 3	mouth	Mund	no
critical	2	(7 * 2) - 6 = 8	piano	Klavier	yes
critical	2	(9 / 3) - 1 = 6	tree	Baum	no
critical	3	(20 / 2) - 9 = 1	foot	Fuß	yes
critical	3	(8 / 8) + 6 = 3	rain	Regen	no
critical	3	(14 * 1) - 8 = 6	group	Gruppe	yes
critical	3	(14 / 7) + 2 = 4	clock	Uhr	yes
critical	3	(6 * 1) + 2 = 3	dust	Staub	no
critical	3	(2 * 2) + 5 = 9	island	Insel	yes
critical	4	(10 / 5) + 1 = 3	dinner	Essen	yes
critical	3	(5 * 2) - 5 = 9	bottle	Flasche	no
critical	3	(10 / 5) + 3 = 9	hill	Hügel	no
critical	4	(12 / 2) - 4 = 2	lake	See	yes
critical	4	(12 / 2) - 5 = 6	king	König	no
critical	4	(6 * 2) - 8 = 9	girl	Frau	no

Appendix D

Critical Stimuli in Experiment 1

English Word		Length	Log frequency[a]	German Translation		Translation Condition					
						CI		II		IC	
THE	HOLE	4	1.76	das	Loch	DAS	OBST	DER	WITZ	DER	LOCH
THE	WOOD	4	1.87	das	Holz	DAS	OBST	DER	WITZ	DER	HOLZ
THE	BREAD	5	1.88	das	Brot	DAS	OBST	DER	WITZ	DER	BROT
THE	ROPE	4	1.51	das	Seil	DAS	MEER	DER	MORD	DER	SEIL
THE	TENT	4	1.58	das	Zelt	DAS	MEER	DER	MORD	DER	ZELT
THE	GOAL	4	1.49	das	Ziel	DAS	MEER	DER	MORD	DER	ZIEL
THE	SHIRT	5	1.67	das	Hemd	DAS	GEMÜSE	DER	FUCHS	DER	HEMD
THE	SHEEP	5	1.61	das	Schaf	DAS	GEMÜSE	DER	FUCHS	DER	SCHAF
THE	PALACE	6	1.64	das	Schloss	DAS	GEMÜSE	DER	FUCHS	DER	SCHLOSS
THE	SCARF	5	0.95	der	Schal	DER	KNOPF	DAS	BRETT	DAS	SCHAL
THE	RABBIT	6	1.07	der	Hase	DER	KNOPF	DAS	BRETT	DAS	HASE
THE	KNIGHT	6	1.09	der	Ritter	DER	KNOPF	DAS	BRETT	DAS	RITTER
THE	SPOON	5	1.14	der	Löffel	DER	SCHIRM	DAS	SCHACH	DAS	LÖFFEL
THE	SCREAM	6	1.17	der	Schrei	DER	SCHIRM	DAS	SCHACH	DAS	SCHREI
THE	COUGH	5	1.14	der	Husten	DER	SCHIRM	DAS	SCHACH	DAS	HUSTEN
THE	TIP	3	1.4	der	Hinweis	DER	DAUMEN	DAS	GEMÄLDE	DAS	HINWEIS
THE	FOAM	4	0.89	der	Schaum	DER	DAUMEN	DAS	GEMÄLDE	DAS	SCHAUM
THE	BELT	4	1.35	der	Gürtel	DER	DAUMEN	DAS	GEMÄLDE	DAS	GÜRTEL
THE	BACK	4	3.09	der	Rücken	DER	WERT	DIE	GEFAHR	DIE	RÜCKEN
THE	HEAD	4	2.66	der	Kopf	DER	WERT	DIE	GEFAHR	DIE	KOPF
THE	LOOK	4	2.77	der	Blick	DER	WERT	DIE	GEFAHR	DIE	BLICK
THE	ENTRANCE	8	1.68	der	Eingang	DER	VORTRAG	DIE	AHNUNG	DIE	EINGANG
THE	ACCIDENT	8	1.7	der	Unfall	DER	VORTRAG	DIE	AHNUNG	DIE	UNFALL
THE	STOMACH	7	1.61	der	Magen	DER	VORTRAG	DIE	AHNUNG	DIE	MAGEN
THE	BONE	4	1.45	der	Knochen	DER	TEPPICH	DAS	ERLEBNIS	DAS	KNOCHEN
THE	PROOF	5	1.5	der	Beweis	DER	TEPPICH	DAS	ERLEBNIS	DAS	BEWEIS
THE	SWEAT	5	1.48	der	Schweiß	DER	TEPPICH	DAS	ERLEBNIS	DAS	SCHWEISS

Appendix E

75 Critical Items Used in Gender Assignment

Type	Word	Gender	Length	Frequency[a]	Translation	Length	Log Frequency[b]
Typical	ABGASE	F	6	14	exhaust	7	0.78
Typical	AUFGABE	F	7	8	assignment	10	0.94
Typical	AUSNAHME	F	8	10	exception	9	1.43
Typical	FAHNE	F	5	12	flag	4	1.32
Typical	FLASCHE	F	7	11	bottle	6	1.92
Typical	GESCHICHTE	F	10	7	story	5	2.23
Typical	HERBERGE	F	8	14	hostel	6	0.77
Typical	HOSE	F	4	12	pants	5	1.22
Typical	KANNE	F	5	16	can	3	3.32
Typical	KARTE	F	5	9	card	4	1.67
Typical	KIRCHE	F	6	8	church	6	2.2
Typical	KISTE	F	5	12	crate	5	0.55
Typical	LIEBE	F	5	9	love	4	2.55
Typical	LÜGE	F	4	12	lie	3	1.92
Typical	MIETE	F	5	11	rent	4	1.64
Typical	MÜNZE	F	5	13	coin	4	0.95
Typical	PUPPE	F	5	10	doll	4	1.27
Typical	REISE	F	5	9	trip	4	1.76
Typical	SCHLANGE	F	8	11	snake	5	1.19
Typical	SCHWÄCHE	F	8	11	weakness	8	1.35
Typical	SPRACHE	F	7	9	language	8	2.12
Typical	STELLE	F	6	8	position	8	2.29
Typical	STIMME	F	6	9	voice	5	2.37
Typical	TÜTE	F	4	13	bag	3	1.8
Typical	WÄSCHE	F	6	13	laundry	7	1.09
Atypical	FERIEN	F	6	11	vacation	8	1.13
Atypical	UMWELT	F	6	10	environment	11	0
Atypical	ZUVERSICHT	F	10	11	confidence	10	1.81
Atypical	FRUCHT	F	6	14	fruit	5	1.77
Atypical	STADT	F	5	6	city	4	2.34
Atypical	ARBEIT	F	6	7	work	4	2.92
Atypical	LEITUNG	F	7	10	pipeline	8	0.66
Atypical	TEE	M	3	12	tea	3	1.95
Atypical	LUFT	F	4	8	air	3	2.4
Atypical	FAHRT	F	5	9	ride	4	1.55
Atypical	MUSIK	F	5	8	music	5	2.13
Atypical	KARTOFFEL	F	9	14	potato	6	1.1
Atypical	NACHT	F	5	7	night	5	2.63

Atypical	AHNUNG	F	6	11	idea	4	2.44
Atypical	ZUKUNFT	F	7	7	future	6	0.92
Atypical	GABEL	F	5	14	fork	4	1.17
Atypical	MUTTER	F	6	8	mother	6	2.63
Atypical	GEBURT	F	6	10	birth	5	1.78
Atypical	ZWIEBEL	F	7	14	onion	5	1.02
Atypical	ANKUNFT	F	7	11	arrival	7	1.58
Atypical	WELT	F	4	6	world	5	2.87
Atypical	ERFAHRUNG	F	9	9	experience	10	2.31
Atypical	FENSTER	F	7	9	window	6	2.13
Atypical	HAUT	F	4	10	skin	4	1.96
Atypical	HOCHZEIT	F	8	11	wedding	7	1.51
Ambiguous	ELEND	N	5	12	misery	6	1.32
Ambiguous	GESPRÄCH	N	8	8	conversation	12	0
Ambiguous	ERLEBNIS	N	8	11	experience	10	2.31
Ambiguous	GÜRTEL	M	6	13	belt	4	1.35
Ambiguous	VEREIN	M	6	8	club	4	1.95
Ambiguous	SCHNITT	M	7	9	cut	3	2.25
Ambiguous	GESCHIRR	N	8	14	dishes	6	1.32
Ambiguous	OHR	N	3	11	ear	3	1.63
Ambiguous	BERUF	M	5	10	job	3	2.39
Ambiguous	KOCH	M	4	9	cook	4	1.7
Ambiguous	BRIEF	M	5	9	letter	6	2.09
Ambiguous	GESPENST	N	8	14	ghost	5	1.31
Ambiguous	KÖRPER	M	6	9	body	4	2.47
Ambiguous	SCHUSS	M	6	13	shot	4	1.95
Ambiguous	BEWEIS	M	6	10	proof	5	1.5
Ambiguous	SCHAL	M	5	13	scarf	5	0.95
Ambiguous	GARTEN	M	6	10	garden	6	2.05
Ambiguous	TRAUM	M	5	10	dream	5	1.77
Ambiguous	KLAVIER	N	7	12	piano	5	1.43
Ambiguous	EINGANG	M	7	11	entrance	8	1.68
Ambiguous	TANZ	M	4	11	dance	5	1.67
Ambiguous	SCHUTZ	M	6	9	protection	10	1.74
Ambiguous	WINTER	M	6	9	winter	6	1.91
Ambiguous	SACK	M	4	12	bag	3	1.8
Ambiguous	RITTER	M	6	12	knight	6	1.09

Notes . [a]Quasthoff, U. (2002); [b]Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995)

Appendix F

Critical Items Used in Experiment 2

English Word	Log Frequency[a]	Length	Gender	Article	German Word	Log Frequency[a]	Frequency[b]	Length	CI		II		IC	
SATURDAY	0	8	M	DER	SAMSTAG	1.60	7	7	DER	WALD	DIE	HERBERGE	DAS	SAMSTAG
FOREST	0	7	M	DER	WALD	1.71	10	4	DER	TROCKNER	DAS	GESPRÄCH	DIE	WALD
DRIER	0.6	5	M	DER	TROCKNER		15	8	DER	SAMSTAG	DAS	BENZIN	DAS	TROCKNER
CONVERSATION	0	12	N	DAS	GESPRÄCH	2.22	8	8	DAS	BENZIN	DER	TROCKNER	DER	GESPRÄCH
GASOLINE	0.69	8	N	DAS	BENZIN	1.23	11	6	DAS	GESCHIRR	DIE	UMWELT	DER	BENZIN
DISHES	1.32	6	N	DAS	GESCHIRR	0.85	14	8	DAS	GESPRÄCH	DIE	LEITUNG	DIE	GESCHIRR
ENVIRONMENT	0	11	F	DIE	UMWELT	1.54	10	6	DIE	LEITUNG	DAS	GESCHIRR	DAS	UMWELT
PIPELINE	0.66	8	F	DIE	LEITUNG	2.03	10	7	DIE	HERBERGE	DER	SAMSTAG	DAS	LEITUNG
HOSTEL	0.77	6	F	DIE	HERBERGE	0.00	14	8	DIE	UMWELT	DER	WALD	DAS	HERBERGE
TRASH	0.67	5	M	DER	MÜLL	0.48	11	4	DER	SENF	DIE	ABGASE	DAS	MÜLL
MUSTARD	0.76	7	M	DER	SENF	0.60	14	4	DER	SCHAL	DAS	RAD	DIE	SENF
SCARF	0.95	5	M	DER	SCHAL	0.00	13	5	DER	MÜLL	DAS	GEPÄCK	DIE	SCHAL
BIKE	0.98	4	N	DAS	RAD	1.40	11	3	DAS	GEPÄCK	DER	SCHAL	DER	RAD
BAGGAGE	0.98	7	N	DAS	GEPÄCK	1.04	21	7	DAS	TUCH	DIE	SCHEIBE	DER	GEPÄCK
CLOTH	1.67	5	N	DAS	TUCH	0.48	13	4	DAS	RAD	DIE	DUSCHE	DIE	TUCH
PANE	0.44	4	F	DIE	SCHEIBE	1.00	12	7	DIE	DUSCHE	DAS	TUCH	DAS	SCHEIBE
SHOWER	1.26	6	F	DIE	DUSCHE	1.00	13	6	DIE	ABGASE	DER	MÜLL	DAS	DUSCHE
EXHAUST	0.78	7	F	DIE	ABGASE	0.00	14	6	DIE	SCHEIBE	DER	SENF	DAS	ABGASE
ARROW	0.96	5	M	DER	PFEIL	0.85	14	5	DER	STIEFEL	DIE	WÄSCHE	DIE	PFEIL
BOOT	1.03	4	M	DER	STIEFEL	0.95	13	7	DER	RITTER	DAS	LINEAL	DAS	STIEFEL
KNIGHT	1.09	6	M	DER	RITTER	0.90	12	6	DER	PFEIL	DAS	GESPENST	DAS	RITTER
RULER	0.94	5	N	DAS	LINEAL	0.00	16	6	DAS	GESPENST	DER	RITTER	DER	LINEAL
GHOST	1.31	5	N	DAS	GESPENST	0.70	14	8	DAS	GEWICHT	DIE	KREIDE	DER	GESPENST
WEIGHT	1.97	6	N	DAS	GEWICHT	1.61	10	7	DAS	LINEAL	DIE	ZITRONE	DER	GEWICHT
CHALK	1.03	5	F	DIE	KREIDE	0.30	14	7	DIE	ZITRONE	DAS	GEWICHT	DAS	KREIDE
LEMON	1.15	5	F	DIE	ZITRONE	0.00	15	7	DIE	WÄSCHE	DER	PFEIL	DAS	ZITRONE
LAUNDRY	1.09	7	F	DIE	WÄSCHE	1.04	13	6	DIE	KREIDE	DER	STIEFEL	DAS	WÄSCHE
CAGE	1.15	4	M	DER	KÄFIG	0.48	13	5	DER	EIMER	DIE	FERIEN	DAS	KÄFIG
BUCKET	1.15	6	M	DER	EIMER	0.70	13	5	DER	BACH	DAS	FASS	DAS	EIMER
CREEK	1.16	5	M	DER	BACH	0.90	11	4	DER	KÄFIG	DAS	LOCH	DAS	BACH

BARREL	1.18	6	N	DAS	FASS	0.48	12	4	DAS	LOCH	DER	BACH	DIE	FASS
HOLE	1.76	4	N	DAS	LOCH	0.95	11	4	DAS	ELEND	DIE	ZWIEBEL	DER	LOCH
MISERY	1.32	6	N	DAS	ELEND	1.15	12	5	DAS	FASS	DIE	KISTE	DER	ELEND
ONION	1.02	5	F	DIE	ZWIEBEL	0.60	14	7	DIE	KISTE	DAS	ELEND	DAS	ZWIEBEL
CRATE	0.55	5	F	DIE	KISTE	1.08	12	5	DIE	FERIEN	DER	KÄFIG	DAS	KISTE
VACATION	1.13	8	F	DIE	FERIEN	1.32	11	6	DIE	ZWIEBEL	DER	EIMER	DAS	FERIEN
FOX	1.17	3	M	DER	FUCHS	0.60	11	5	DER	AUFSATZ	DIE	MÜNZE	DAS	FUCHS
ESSAY	1.18	5	M	DER	AUFSATZ	1.00	14	7	DER	KNOPF	DAS	MEHL	DIE	AUFSATZ
BUTTON	1.25	6	M	DER	KNOPF	0.78	13	5	DER	FUCHS	DAS	SCHACH	DAS	KNOPF
FLOUR	1.43	5	N	DAS	MEHL	0.60	9	4	DAS	SCHACH	DER	KNOPF	DIE	MEHL
CHESS	1.2	5	N	DAS	SCHACH	1.00	13	6	DAS	BLATT	DIE	PUPPE	DIE	SCHACH
LEAF	1.22	4	N	DAS	BLATT	1.66	9	5	DAS	MEHL	DIE	SEIDE	DER	BLATT
DOLL	1.27	4	F	DIE	PUPPE	0.70	10	5	DIE	SEIDE	DAS	BLATT	DAS	PUPPE
SILK	1.41	4	F	DIE	SEIDE	0.78	12	5	DIE	MÜNZE	DER	FUCHS	DAS	SEIDE
COIN	0.95	4	F	DIE	MÜNZE	0.90	13	5	DIE	PUPPE	DER	AUFSATZ	DAS	MÜNZE
TRUNK	1.32	5	M	DER	STAMM	1.23	12	5	DER	GÜRTEL	DIE	WOLKE	DAS	STAMM
BELT	1.35	4	M	DER	GÜRTEL	0.78	13	6	DER	BART	DAS	SCHWEIN	DAS	GÜRTEL
BEARD	1.37	5	M	DER	BART	1.00	12	4	DER	STAMM	DAS	SEIL	DIE	BART
PIG	1.28	3	N	DAS	SCHWEIN	0.70	13	7	DAS	SEIL	DER	BART	DER	SCHWEIN
ROPE	1.51	4	N	DAS	SEIL	0.78	13	4	DAS	KORN	DIE	LEBER	DIE	SEIL
SEED	1.46	4	N	DAS	KORN	0.78	12	4	DAS	SCHWEIN	DIE	GABEL	DER	KORN
LIVER	1.16	5	F	DIE	LEBER	0.90	13	5	DIE	GABEL	DAS	KORN	DAS	LEBER
FORK	1.17	4	F	DIE	GABEL	0.60	14	5	DIE	WOLKE	DER	STAMM	DAS	GABEL
CLOUD	1.51	5	F	DIE	WOLKE	0.70	13	5	DIE	LEBER	DER	GÜRTEL	DAS	WOLKE
TIP	1.4	3	M	DER	HINWEIS	1.71	10	7	DER	TOPF	DIE	FREUDE	DIE	HINWEIS
POT	1.41	3	M	DER	TOPF	0.85	12	4	DER	RAHMEN	DAS	GEFÄNGNIS	DIE	TOPF
FRAME	1.44	5	M	DER	RAHMEN	2.06	8	6	DER	HINWEIS	DAS	EI	DAS	RAHMEN
PRISON	1.85	6	N	DAS	GEFÄNGNIS	1.67	9	9	DAS	EI	DER	RAHMEN	DIE	GEFÄNGNIS
EGG	1.58	3	N	DAS	EI	0.95	12	2	DAS	GLÜCK	DIE	ZUKUNFT	DER	EI
LUCK	1.67	4	N	DAS	GLÜCK	1.97	8	5	DAS	GEFÄNGNIS	DIE	WURST	DER	GLÜCK
APOLOGY	0.92	7	F	DIE	ZUKUNFT	2.29	7	7	DIE	WURST	DAS	GLÜCK	DAS	ZUKUNFT
SAUSAGE	0.83	7	F	DIE	WURST	0.95	13	5	DIE	FREUDE	DER	HINWEIS	DAS	WURST
JOY	1.62	3	F	DIE	FREUDE	1.90	9	6	DIE	ZUKUNFT	DER	TOPF	DAS	FREUDE
PROOF	1.5	5	M	DER	BEWEIS	1.68	10	6	DER	FLÜGEL	DIE	LÜGE	DIE	BEWEIS
WING	1.53	4	M	DER	FLÜGEL	1.34	11	6	DER	WITZ	DAS	SCHICKSAL	DAS	FLÜGEL
JOKE	1.55	4	M	DER	WITZ	1.04	11	4	DER	BEWEIS	DAS	TOR	DIE	WITZ
FATE	1.53	4	N	DAS	SCHICKSAL	1.90	10	9	DAS	TOR	DER	WITZ	DER	SCHICKSAL
GATE	1.71	4	N	DAS	TOR	1.46	8	3	DAS	BROT	DIE	JUGEND	DER	TOR
BREAD	1.88	5	N	DAS	BROT	1.45	11	4	DAS	SCHICKSAL	DIE	FAHRT	DER	BROT
YOUTH	1.82	5	F	DIE	JUGEND	2.16	10	6	DIE	FAHRT	DAS	BROT	DAS	JUGEND
RIDE	1.55	4	F	DIE	FAHRT	1.76	9	5	DIE	LÜGE	DER	BEWEIS	DAS	FAHRT
LIE	1.92	3	F	DIE	LÜGE	1.38	12	4	DIE	JUGEND	DER	FLÜGEL	DAS	LÜGE
GIANT	1.56	5	M	DER	RIESE	0.00	13	5	DER	SPIEGEL	DIE	KETTE	DIE	RIESE
MIRROR	1.62	6	M	DER	SPIEGEL	1.56	9	7	DER	STAUB	DAS	SCHAF	DAS	SPIEGEL
DUST	1.67	4	M	DER	STAUB	1.28	11	5	DER	RIESE	DAS	SCHLOSS	DIE	STAUB

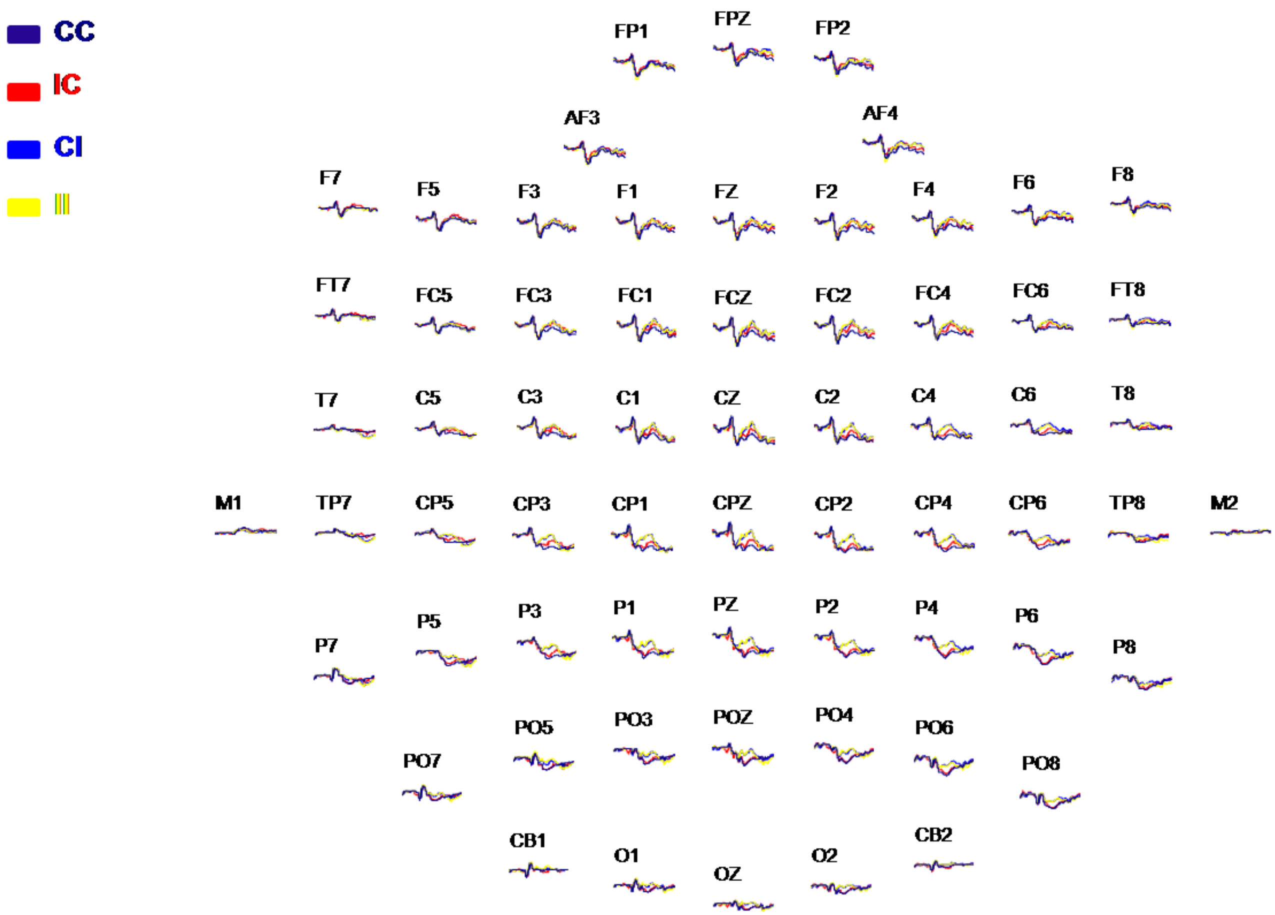
SHEEP	1.61	5	N	DAS	SCHAF	0.30	13	5	DAS	SCHLOSS	DER	STAUB	DER	SCHAF
PALACE	1.64	6	N	DAS	SCHLOSS	1.81	10	7	DAS	MESSER	DIE	GLOCKE	DIE	SCHLOSS
KNIFE	1.58	5	N	DAS	MESSER	0.85	11	6	DAS	SCHAF	DIE	WAFFE	DER	MESSER
BELL	1.61	4	F	DIE	GLOCKE	0.00	13	6	DIE	WAFFE	DAS	MESSER	DAS	GLOCKE
WEAPON	1.4	6	F	DIE	WAFFE	1.38	11	5	DIE	KETTE	DER	RIESE	DAS	WAFFE
CHAIN	1.53	5	F	DIE	KETTE	1.23	12	5	DIE	GLOCKE	DER	SPIEGEL	DAS	KETTE
DANCE	1.67	5	M	DER	TANZ	1.32	11	4	DER	UNFALL	DIE	ANKUNFT	DIE	TANZ
ACCIDENT	1.7	8	M	DER	UNFALL	1.48	9	6	DER	MORD	DAS	OHR	DIE	UNFALL
MURDER	1.71	6	M	DER	MORD	1.49	10	4	DER	TANZ	DAS	KLAVIER	DIE	MORD
EAR	1.63	3	N	DAS	OHR	1.40	11	3	DAS	KLAVIER	DER	MORD	DER	OHR
PIANO	1.43	5	N	DAS	KLAVIER	1.15	12	7	DAS	MEER	DIE	AUFGABE	DER	KLAVIER
OCEAN	1.4	5	N	DAS	MEER	1.75	10	4	DAS	OHR	DIE	HOCHZEIT	DER	MEER
ASSIGNMENT	0.94	10	F	DIE	AUFGABE	2.33	8	7	DIE	HOCHZEIT	DAS	MEER	DAS	AUFGABE
WEDDING	1.51	7	F	DIE	HOCHZEIT	1.26	11	8	DIE	ANKUNFT	DER	TANZ	DAS	HOCHZEIT
ARRIVAL	1.58	7	F	DIE	ANKUNFT	1.30	11	7	DIE	AUFGABE	DER	UNFALL	DAS	ANKUNFT
PATH	1.71	4	M	DER	WEG	2.44	7	3	DER	MANTEL	DIE	GEFAHR	DIE	WEG
COAT	1.73	4	M	DER	MANTEL	1.30	12	6	DER	MOND	DAS	ZIEL	DIE	MANTEL
MOON	1.73	4	M	DER	MOND	1.77	11	4	DER	WEG	DAS	HEMD	DIE	MOND
GOAL	1.49	4	N	DAS	ZIEL	2.27	7	4	DAS	HEMD	DER	MOND	DER	ZIEL
SHIRT	1.67	5	N	DAS	HEMD	1.23	12	4	DAS	SCHIFF	DIE	MIETE	DER	HEMD
SHIP	1.67	4	N	DAS	SCHIFF	1.72	10	6	DAS	ZIEL	DIE	FAHNE	DER	SCHIFF
RENT	1.64	4	F	DIE	MIETE	1.36	11	5	DIE	FAHNE	DAS	SCHIFF	DAS	MIETE
FLAG	1.32	4	F	DIE	FAHNE	1.08	12	5	DIE	GEFAHR	DER	WEG	DAS	FAHNE
DANGER	1.87	6	F	DIE	GEFAHR	2.17	8	6	DIE	MIETE	DER	MANTEL	DAS	GEFAHR
FLIGHT	1.76	6	M	DER	FLUG	1.32	10	4	DER	TRAUM	DIE	TÜTE	DIE	FLUG
DREAM	1.77	5	M	DER	TRAUM	1.30	10	5	DER	RAUCH	DAS	TIER	DIE	TRAUM
SMOKE	1.77	5	M	DER	RAUCH	1.11	11	5	DER	FLUG	DAS	LIED	DIE	RAUCH
ANIMAL	2.06	6	N	DAS	TIER	1.57	10	4	DAS	LIED	DER	RAUCH	DER	TIER
SONG	1.53	4	N	DAS	LIED	1.36	11	4	DAS	KLEID	DIE	ADRESSE	DER	LIED
DRESS	1.93	5	N	DAS	KLEID	1.18	12	5	DAS	TIER	DIE	REISE	DER	KLEID
ADDRESS	1.7	7	F	DIE	ADRESSE	1.36	11	7	DIE	REISE	DAS	KLEID	DAS	ADRESSE
TRIP	1.76	4	F	DIE	REISE	1.48	9	5	DIE	TÜTE	DER	FLUG	DAS	REISE
BAG	1.8	3	F	DIE	TÜTE	0.70	13	4	DIE	ADRESSE	DER	TRAUM	DAS	TÜTE
WASTE	1.81	5	M	DER	ABFALL	0.78	13	6	DER	LÄRM	DIE	HAUT	DIE	ABFALL
NOISE	1.81	5	M	DER	LÄRM	1.28	11	4	DER	REGEN	DAS	FETT	DIE	LÄRM
RAIN	1.86	4	M	DER	REGEN	1.63	10	5	DER	ABFALL	DAS	FLEISCH	DAS	REGEN
FAT	1.94	3	N	DAS	FETT	1.11	12	4	DAS	FLEISCH	DER	REGEN	DER	FETT
MEAT	1.86	4	N	DAS	FLEISCH	1.52	10	7	DAS	HOLZ	DIE	DIÄT	DER	FLEISCH
WOOD	1.87	4	N	DAS	HOLZ	1.60	11	4	DAS	FETT	DIE	UHR	DER	HOLZ
DIET	1.75	4	F	DIE	DIÄT	0.60	13	4	DIE	UHR	DAS	HOLZ	DAS	DIÄT
CLOCK	1.56	5	F	DIE	UHR	2.86	6	3	DIE	HAUT	DER	ABFALL	DAS	UHR
SKIN	1.96	4	F	DIE	HAUT	1.60	10	4	DIE	DIÄT	DER	LÄRM	DAS	HAUT
TREE	1.86	4	M	DER	BAUM	1.38	10	4	DER	HALS	DIE	GEBURT	DIE	BAUM
NECK	1.88	4	M	DER	HALS	1.57	11	4	DER	SCHMERZ	DAS	DACH	DIE	HALS
PAIN	1.88	4	M	DER	SCHMERZ	1.20	11	8	DER	BAUM	DAS	GEWEHR	DIE	SCHMERZ
ROOF	1.68	4	N	DAS	DACH	1.40	9	4	DAS	GEWEHR	DER	SCHMERZ	DER	DACH
GUN	1.81	3	N	DAS	GEWEHR	1.08	13	6	DAS	KREUZ	DIE	FLASCHE	DER	GEWEHR
CROSS	1.82	5	N	DAS	KREUZ	1.46	10	5	DAS	DACH	DIE	BRÜCKE	DIE	KREUZ

BOTTLE	1.92	6	F	DIE	FLASCHE	1.42	11	7	DIE	BRÜCKE	DAS	KREUZ	DAS	FLASCHE
BRIDGE	1.79	6	F	DIE	BRÜCKE	0.00	10	6	DIE	GEBURT	DER	BAUM	DAS	BRÜCKE
BIRTH	1.78	5	F	DIE	GEBURT	1.36	10	6	DIE	FLASCHE	DER	HALS	DAS	GEBURT
SUCCESS	2.02	7	M	DER	ERFOLG	2.27	7	6	DER	DRUCK	DIE	SPRACHE	DIE	ERFOLG
PRESSURE	2.03	8	M	DER	DRUCK	1.70	8	5	DER	SCHLAF	DAS	DORF	DIE	DRUCK
SLEEP	2.09	5	M	DER	SCHLAF	1.36	11	5	DER	ERFOLG	DAS	ERLEBNIS	DIE	SCHLAF
VILLAGE	2.13	7	N	DAS	DORF	1.72	10	4	DAS	ERLEBNIS	DER	SCHLAF	DER	DORF
EXPERIENCE	2.31	10	N	DAS	ERLEBNIS	1.20	11	8	DAS	SCHILD	DIE	EHE	DER	ERLEBNIS
SIGN	2.01	4	N	DAS	SCHILD	0.70	12	6	DAS	DORF	DIE	STELLE	DER	SCHILD
MARRIAGE	1.98	8	F	DIE	EHE	1.81	10	3	DIE	STELLE	DAS	SCHILD	DAS	EHE
POSITION	2.29	8	F	DIE	STELLE	2.30	8	6	DIE	SPRACHE	DER	ERFOLG	DAS	STELLE
LANGUAGE	2.12	8	F	DIE	SPRACHE	2.12	9	7	DIE	EHE	DER	DRUCK	DAS	SPRACHE
LETTER	2.09	6	M	DER	BRIEF	1.93	9	5	DER	RAUM	DIE	BEWEGUNG	DIE	BRIEF
SPACE	2.1	5	M	DER	RAUM	2.20	8	4	DER	BERICHT	DAS	BLUT	DIE	RAUM
REPORT	2.13	6	M	DER	BERICHT	2.11	8	7	DER	BRIEF	DAS	BILD	DIE	BERICHT
BLOOD	2.15	5	N	DAS	BLUT	1.60	10	4	DAS	BILD	DER	BERICHT	DER	BLUT
PICTURE	2.03	7	N	DAS	BILD	2.35	8	4	DAS	GESETZ	DIE	RUHE	DER	BILD
LAW	2.22	3	N	DAS	GESETZ	1.97	8	6	DAS	BLUT	DIE	WAHL	DER	GESETZ
REST	2.34	4	F	DIE	RUHE	1.87	9	4	DIE	WAHL	DAS	GESETZ	DAS	RUHE
ELECTION	1.86	8	F	DIE	WAHL	2.03	7	4	DIE	BEWEGUNG	DER	BRIEF	DAS	WAHL
MOVEMENT	2.28	8	F	DIE	BEWEGUNG	2.04	9	8	DIE	RUHE	DER	RAUM	DAS	BEWEGUNG
FLOOR	2.21	5	M	DER	BODEN	2.16	8	5	DER	ABEND	DIE	FRAGE	DAS	BODEN
EVENING	2.27	7	M	DER	ABEND	2.05	7	5	DER	GRUND	DAS	HERZ	DIE	ABEND
REASON	2.37	6	M	DER	GRUND	2.38	7	5	DER	BODEN	DAS	SPIEL	DIE	GRUND
HEART	2.16	5	N	DAS	HERZ	1.90	9	4	DAS	SPIEL	DER	GRUND	DER	HERZ
GAME	2.17	4	N	DAS	SPIEL	2.16	7	5	DAS	GEFÜHL	DIE	STRASSE	DER	SPIEL
FEELING	2.31	7	N	DAS	GEFÜHL	1.85	9	6	DAS	HERZ	DIE	AHNUNG	DER	GEFÜHL
STREET	2.41	6	F	DIE	STRASSE	2.31	7	6	DIE	AHNUNG	DAS	GEFÜHL	DAS	STRASSE
IDEA	2.44	4	F	DIE	AHNUNG	1.32	11	6	DIE	FRAGE	DER	BODEN	DAS	AHNUNG
QUESTION	2.47	8	F	DIE	FRAGE	2.80	7	5	DIE	STRASSE	DER	ABEND	DAS	FRAGE
JOB	2.39	3	M	DER	BERUF	1.79	10	5	DER	KÖRPER	DIE	HILFE	DIE	BERUF
BODY	2.47	4	M	DER	KÖRPER	1.78	9	6	DER	PUNKT	DAS	BUCH	DAS	KÖRPER
POINT	2.61	5	M	DER	PUNKT	2.04	9	5	DER	BERUF	DAS	GEBIET	DIE	PUNKT
BOOK	2.44	4	N	DAS	BUCH	2.00	8	4	DAS	GEBIET	DER	PUNKT	DER	BUCH
AREA	2.31	4	N	DAS	GEBIET	2.22	9	6	DAS	GELD	DIE	LUFT	DER	GEBIET
MONEY	2.61	5	N	DAS	GELD	2.26	7	4	DAS	BUCH	DIE	STIMME	DER	GELD
AIR	2.4	3	F	DIE	LUFT	2.02	8	4	DIE	STIMME	DAS	GELD	DAS	LUFT
VOICE	2.37	5	F	DIE	STIMME	2.16	9	6	DIE	HILFE	DER	BERUF	DAS	STIMME
HELP	2.58	4	F	DIE	HILFE	2.33	8	5	DIE	LUFT	DER	KÖRPER	DAS	HILFE

Notes. [a]Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995); [b]Quasthoff, U. (2002)

Appendix G

ERP Waveforms of all 64 Channels



Appendix H

Critical Items Used in Experiment 3

Condition	Gender	German Compound	L	F C1[a]	L C1	F C2[a]	L C2	English Translation	L	L C1	Log F[b]	L C2	Log F[b]
AA	FF	ARBEITSSTELLE	13	7	6	8	6	WORKPLACE	9	4	2.92	5	2.74
AA	MM	BAUCHTANZ	9	11	5	11	4	BELLY DANCE	10	5	1.26	5	1.67
AA	MM	BESENSTIEL	10	14	5	15	5	BROOMSTICK	10	5	0.87	5	1.75
AA	NN	BETTTUCH	8	10	4	13	4	BED SHEET	8	3	2.39	5	1.55
VA	NM	BIERGARTEN	10	10	4	10	6	BEER GARDEN	10	4	1.68	6	2.05
VA	NM	BLUTDRUCK	9	10	4	8	5	BLOOD PRESSURE	13	5	2.15	8	2.03
AA	MM	BRIEFKASTEN	11	9	5	12	6	MAILBOX	7	4	1.42	3	1.9
VA	MF	BUSFAHRT	8	10	3	9	5	BUS RIDE	7	3	1.82	4	1.55
AA	FF	FENSTERSCHEIBE	14	9	7	12	7	WINDOW PANE	10	6	2.13	4	0.44
AA	FF	FERIENWOHNUNG	13	11	6	8	7	VACATION APARTMENT	17	8	1.13	9	1.65
AA	MM	FILMSTAR	8	8	4	10	4	MOVIE STAR	9	5	1.49	4	1.73
AA	MM	FINGERNAGEL	11	10	6	11	5	FINGERNAIL	10	6	1.7	4	1.11
AA	MM	FLOHMARKT	9	13	4	7	5	FLEA MARKET	10	4	0.43	6	2.13
VA	FM	FRUCHTSAFT	10	14	6	13	4	FRUIT JUICE	10	5	1.77	5	1.33
AA	MM	FUSSBALL	7	9	3	9	4	FOOTBALL	8	4	2.07	4	1.97
VA	FM	GEBURTSORT	9	10	6	8	3	BIRTHPLACE	10	5	1.78	5	2.74
VA	NM	GESCHIRRSPÜLER	14	14	8	20	6	DISHWASHER	10	4	1.36	6	0.43
VA	NM	GOLDFISCH	9	10	4	11	5	GOLDFISH	8	4	1.95	4	2.24
VA	NM	HAARSCHNITT	11	11	4	9	7	HAIRCUT	7	4	2.28	3	2.25
VA	NF	HAUSAUFGABE	11	7	4	8	7	HOMEWORK	8	4	2.69	4	2.92
VA	FN	HOCHZEITSKLEID	14	11	8	12	5	WEDDING DRESS	12	7	1.51	5	1.93
AA	FF	JUGENDHERBERGE	14	10	6	14	8	YOUTH HOSTEL	11	5	1.82	6	0.77
VA	FM	KARTOFFELSALAT	14	14	9	13	5	POTATO SALAD	11	6	1.1	5	1.22
AA	MM	KÄSEKUCHEN	10	12	4	12	6	CHEESECAKE	10	6	1.46	4	1.35

VA	NF	KLAVIERMUSIK	12	12	7	8	5	PIANO MUSIC	10	5	1.43	5	2.13
VA	MN	KOCHBUCH	8	9	4	8	4	COOKBOOK	8	4	1.7	4	2.44
AA	MM	KÖRPERTEIL	10	9	6	7	4	BODY PART	8	4	2.47	4	2.7
VA	MF	KREDITKARTE	11	11	6	9	5	CREDIT CARD	10	6	1.62	4	1.67
VA	NF	LEDERHOSE	9	12	5	12	4	LEATHER PANTS	12	7	1.57	5	1.22
VA	FN	LEITUNGSWASSER	14	10	7	8	6	TAP WATER	8	3	1.33	5	2.64
AA	FF	LIEBESGESCHICHTE	16	9	5	7	10	LOVE STORY	9	4	2.55	5	2.23
AA	FF	LUFTVERSCHMUTZUNG	14	8	4	14	13	AIR POLLUTION	12	3	2.4	9	1.51
AA	MM	MÜLLEIMER	9	11	4	13	5	TRASHCAN	8	5	0.67	3	3.32
AA	FF	MUTTERSPRACHE	13	8	6	9	7	MOTHER TONGUE	12	6	2.63	6	1.54
VA	NM	OHRRING	7	11	3	10	4	EARRING	7	3	1.63	4	1.83
VA	NF	PLASTIKTÜTE	11	13	7	13	4	PLASTIC BAG	10	7	1.7	3	1.8
VA	NM	RADWEG	6	11	3	7	3	BIKE PATH	8	4	0.98	4	1.71
AA	MM	REGENWALD	9	10	5	10	4	RAIN FOREST	10	4	1.86	6	1.84
VA	FN	REISEBÜRO	9	9	5	9	4	TRAVEL AGENCY	12	6	1.82	6	1.48
AA	MM	SAMSTAGABEND	12	7	7	7	5	SATURDAY EVENING	15	8		7	2.27
AA	MM	SCHLAFSACK	10	11	6	12	4	SLEEPING BAG	11	8	1.67	3	1.8
AA	MM	SESSELLIFT	10	13	6	14	4	CHAIRLIFT	9	5	2.02	4	1.67
AA	MM	SKISTIEFEL	10	12	3	13	7	SKI BOOT	7	3	0.86	4	1.03
VA	NM	SPIELPLATZ	10	7	5	6	5	PLAYGROUND	10	4	2.44	6	2.28
AA	MM	SPORTVEREIN	11	8	5	8	6	SPORTS CLUB	10	5	1.52	4	1.95
AA	MM	STACHELDRAHT	12	14	7	13	5	BARBED WIRE	10	6	0.96	4	1.56
VA	FN	STADTTOR	8	6	5	8	3	CITY GATE	8	4	2.34	4	1.71
VA	MF	TEEKANNE	8	12	3	16	5	TEAPOT	6	3	1.95	3	1.41
AA	NN	TELEFONGESPRÄCH	15	10	7	8	8	TELEPHONE CALL	13	9	2.01	4	2.38
AA	MM	TRAUMBERUF	10	10	5	10	5	DREAM JOB	8	5	1.77	3	2.39
VA	FM	UMWELTSCHUTZ	12	10	6	9	6	ENVIRONMENTAL PROTECTION	23	13	1.5441	10	1.74
VA	MN	VERKEHRSSCHILD	14	14	7	12	6	TRAFFIC SIGN	11	7	1.74	4	2.01
AA	MM	VOGELKÄFIG	10	11	5	13	5	BIRDCAGE	8	4	1.63	4	1.15
AA	NN	VOLKSLIED	9	9	4	11	4	FOLKSONG	8	4	1.26	4	1.53
VA	FM	WÄSCHETROCKNER	14	13	6	15	8	CLOTHES DRYER	12	7	2.1	5	0.33
VA	MF	WEINFLASCHE	11	10	4	11	7	WINE BOTTLE	10	4	1.87	6	1.92
VA	FM	WELTKRIEG	9	6	4	8	5	WORLD WAR	8	5	2.87	3	2.53
VA	NM	WETTERBERICHT	13	9	6	8	7	WEATHER REPORT	13	7	1.83	6	2.13
AA	MM	WINTERMANTEL	12	9	6	12	6	WINTER COAT	10	6	1.91	4	1.73
VA	MN	ZUGABTEIL	9	9	3	15	6	TRAIN COMPARTMENT	16	5	1.9	11	

Notes. [a]Quasthoff, U. (2002); [b]Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995); AA = agreement, agreement; VA = violation, agreement; L = length; F = frequency; C1 = constituent 1; C2 = constituent 2; Gender: F = feminine, M = masculine, N = neuter

VITA

Susan Christine Bobb

Education

2008	Ph.D. in Psychology, Option in Applied Linguistics, The Pennsylvania State University
2006	M.S. in Psychology, The Pennsylvania State University
2003	M.A. in Linguistics, The University of Virginia
2000	B.S. in Psychology, double major in French, *Summa Cum Laude*, Sweet Briar College

Research Support

2007-2008	National Science Foundation Dissertation Research Grant (with Judith F. Kroll). BCS - 0720155: *Morphology in bilingual language processing* ($11,987)
2007	German Academic Exchange Program (DAAD) short-term research grant
2007	Dissertation Support Grant, College of the Liberal Arts Research and Graduate Studies Office, The Pennsylvania State University (release time)

Selected Publications

Kroll, J. F., Bobb, S. C., & Wodniecka, Z. (2006). Language selectivity is the exception, not the rule: Arguments against a fixed locus of language selection in bilingual speech. *Bilingualism: Language and Cognition 9,* 119-135.

Kroll, J. F., Bobb, S. C., Misra, M., & Guo, T. (2008). Language selection in bilingual speech: Evidence for inhibitory processes. *Acta Psychologica 128* (3), 416-430.

Bobb, S. C., Hoshino, N., & Kroll, J. F. (in press). The role of language cues in constraining cross-language activity. In EUROSLA Yearbook, Volume 8 (Ed., L. Roberts). Amsterdam: John Benjamins Publishing Company.

Jackson, C. N. & Bobb, S. C. (under review). The processing and comprehension of *wh*-questions among L2 German speakers.

Selected Presentations

Bobb, S.C. (2008, March). Morphology in bilingual language processing: The role of second language proficiency in acquiring grammatical gender. Poster presented at the 23rd Annual Graduate Research Exhibit, The Pennsylvania State University, University Park, PA.

Guo, T., Misra, M., Bobb, S.C. & Kroll, J.F. (2007). Behavioral and ERP Measures of Competition in Bilingual Spoken Word Production. Paper read by S.C. Bobb at the International Symposium on Bilingualism (ISB) 6, May 30-June 2, Hamburg, Germany.

Bobb, S.C., & Kroll, J.F. (2006). Priming one of a bilingual's two languages for spoken production: Can language selection be biased by repetition in picture naming? Poster presented at the fifth annual meeting of the Mental Lexicon, October 11-13, Montréal, Canada.

Golombek, P. and Bobb, S.C. (2005, June). Bi-directional learning between teacher and supervisor through journal writing. Paper presented at the Language Teacher Education conference, Minneapolis, MN.

Bobb, S.C., Wodniecka, Z., Kroll, J.F., & Green, D.W. (2005, March). Modulating cross-language activation in a primed language-switching paradigm. Poster presented at the International Symposium on Bilingualism (ISB) 5, Barcelona, Spain.

Selected Awards/Honors

2008	The Pennsylvania State University Alumni Association Dissertation Award
2006	Fellowship, Summer Institute in Cognitive Neuroscience. Dartmouth College, Hanover, NH.