THE PENNSYLVANIA STATE UNIVERSITY
SCHREYER HONORS COLLEGE

PROGRAM IN LINGUISTICS

THE ACQUISITION OF BILINGUAL GENDER IN GERMAN-FRENCH BILINGUALS

LEAH PAPPAS
SPRING 2014

A thesis
submitted in partial fulfillment
of the requirements
for baccalaureate degrees in
Philosophy and in French and Francophone studies
with honors in Linguistics

Reviewed and approved* by the following:

Carrie Jackson
Professor of German and Linguistics
Thesis Supervisor

Lisa Reed
Professor of French and Linguistics
Honors Advisor

Janet Van Hell
Professor of Psychology and Linguistics
Faculty Reader

* Signatures are on file in the Schreyer Honors College.

ABSTRACT

It is commonly known that learning grammatical gender in a second language (L2) is incredibly difficult. Although many studies have investigated L2 gender processing, there is still gray area as to whether or not these difficulties are due to difficulty learning the L2 gender system or first language (L1) influence on L2 gender production. Most studies that have examined how L1 gender influences the learning and production of L2 gender have focused on languages with the same number of gender categories – Czech-German (Bordag & Pechman, 2007), Greek-German (Salamoura & Williams, 2007) – and/or languages that are closely related to one another and, thus, have a large degree of overlap in regards to gender assignment (German-Dutch, Sabourin, Stowe, & de Haan, 2006). Studying German-French learners provides an opportunity to examine L1 gender influence between two different languages with incongruent gender systems. While German has a three-gender system –masculine, feminine, and neuter– French has a two-gender system – masculine and feminine. This study will compare the reaction times in an L2 production task between L1 German - L2 French learners in order to examine how the L1 gender system influences L2 gender learning, especially among nouns with no gender equivalent in the L2.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to give a special thanks to the following that have helped me complete this senior thesis:

PIRE 2013 team and NSF-PIRE grant OISE-0968369, *Bilingualism, Mind and Brain: An interdisciplinary program in cognitive psychology, linguistics, and cognitive neuroscience*

Liberal Arts CEN Funding

Janet Van Hell

Carrie Jackson

Katharina Spalek

Courtney Johnson-Fowler

Sharon Elder

Bobby Marker

Sarah Fairchild

Introduction

Any second language learner of a Romance language or one of the multitudes of other languages that have grammatical gender can understand the woes of mastering the grammatical gender system of the second language (L2) and how difficult it is to obtain native-like mastery of L2 gender. No matter the number of grammatical genders a language may have, it is very difficult for an L2 language learner to master the use of gender in the L2 due to its often arbitrary nature. Although the phonological form of a word often provides cues to its gender, there is no inherent reason that a table should be feminine while a cat should be a masculine. Often a same semantic idea can have different genders in different languages. One example of this is the word "table", which is feminine in French, *la$_F$ table,* and masculine in German, *der$_M$ Tisch*. Even more complicated is the fact that the same semantic idea can have two different genders in the same language, such as "car" in French; *l$_M$'auto* is masculine but *la$_F$ voiture* is feminine. It is possible, though, that certain aspects of an L2 gender may be easier or more difficult to produce based on the gender in one's native language (L1) and the extent to which the L2 gender system overlaps with that of the L1.

But the extent to which the L1 gender system influences the acquisition and processing of L2 gender is still an open question. Previous studies have found evidence that all of a speaker's languages are always active and always compete with each other (Kroll & Tokowicz, 2005). Therefore, an L2 speaker must search the grammar and lexicon of all of his or her languages before choosing the correct word or phrase to produce (De Bot, 1992). This interaction can involve all levels of language: phonology, lexicon, and syntax and may even extend to a lexical-

syntactic structure like grammatical gender. L1 gender may influence the ease and speed at which an L2 speaker produces a gender-marked word or phrase in his or her L2 (e.g., Bordag & Pechman, 2007). If gender is congruent, it may facilitate L2 production, while if gender conflicts between the two languages, it may inhibit production.

A classic study by Costa, Kovacic, Franck, and Carmazza (2003) investigated this topic. They proposed two possible mappings of bilingual gender in the brain. The integrated view suggests that both gender systems are always active and compete during production. The autonomous view states that the two gender systems of a bilingual are kept separate and do not influence each other during production. To test which of these theories is correct, Costa et al. conducted a picture-naming study with three different bilingual populations: Croatian-Italian, Italian-French, and Catalan-Spanish bilinguals, all of whom were highly proficient, early bilinguals. The study included items that had the same (congruent) gender in both languages, like the word for "broom" which in Italian is *scopa*$_{\text{F}}$ and in Croatian is *Metla*$_{F,}$ or items that are different (incongruent) gender in both languages, such as the word for "leaf" which, in Italian, is feminine, *Foglia*$_{F,}$ and in Croatian is *List*$_{M,}$ which is masculine. The results revealed a slight trend towards incongruent words leading to longer naming, or reaction times (RTs), but this difference was not statistically significant. Additionally, monolingual participants showed the same trend as the bilingual participants, which further weakened these findings. Costa et al. (2003) explain these null results by stating that the gender systems of a speaker's languages are independent of each other, which leads to no differences between the production of congruent and incongruent stimuli.

Subsequent studies, however, have found the opposite results. Paolieri, Cubelli, Macizo, Bajo, Lotto, and Job (2010) conducted a recent study building upon the results of Costa et al. In a series of experiments, Paolieri et al. (2010) asked how L1 grammatical gender affects L2 production. Italian-Spanish bilinguals completed a picture-naming task in the first and second

experiments and a translation task in the third experiment. The results showed a significant difference between the participants' reaction times for the congruent and the incongruent items; RTs for congruent items were faster while RTs for incongruent items were longer. These results indicate that the L1 influences L2 production. With congruent items, the gender of the L1 facilitates production of the L2. Likewise, when items are incongruent, the difference in gender inhibits L2 production. From these results, the researchers concluded that in addition to cross-language interaction at the semantic and word-form levels, syntactic information associated with a word's lexical representation, like grammatical gender, is also integrated across a bilingual speaker's two languages.

Bordag and Pechman (2007) obtained results that paralleled Paolieri et al. (2010). Similar to previous studies (e.g., Costa et al., 2003; Paolieri et al., 2010), in Experiment 2 of their 2007 paper, they examined whether L1 gender mappings influence the production of determiner+noun phrases in the L2. Like the previously mentioned studies, they conducted a picture-naming experiment, and participants were completely in the bilingual mode. However, Bordag and Pechman (2007) examined L1 Czech speakers who were late learners of L2 German; both of the speakers' languages have three genders (masculine, feminine, and neuter). Participants were asked to name pictures in their L2 German. The results showed facilitation on items with congruent gender across the two languages and inhibition with incongruent gender items, similar to Paolieri et al. (2010). The researchers concluded that the gender nodes are shared across languages. Therefore, both the L1 and the L2 are activated and compete during the process of selection and production.

The divergent results of the Costa et al. (2003) study versus Paolieri et al. (2010) and Bordag and Pechman (2007) raise many questions regarding L1 interaction in the production of L2 grammatical gender. While Costa et al. (2003) reported evidence that suggests bilingual gender systems are kept separate and do not influence each other, other studies find the opposite

effect, suggesting that the gender systems of a speaker's languages interact during production. This has led other researchers to question whether difficulties in L2 gender production are caused by cross-language interaction or are merely a by-product of difficulty in learning L2 gender.

A study by Lemhöfer, Spalek and Schriefers (2008) elicited results suggesting the latter of the two possibilities, that L2 gender production difficulties may actually be due to difficulties acquiring L2 grammatical gender. This study, like those mentioned above, investigated whether there was L1 gender influence when producing L2 gender among German-Dutch bilinguals. The experiments consisted of a lexical decision task and two picture-naming tasks, one with training before and one without. Although Lemhöfer et al. did find an effect suggesting that the L1 influences L2 production, error rates were so high (close to 60% with gender incongruent nouns), that they raised the possibility of an alternative explanation for their results, in particular, the issue of the difficulty in acquiring L2 gender. They acknowledge that acquiring the correct L2 gender assignment for words that have a different gender in the L1 may be particularly difficult, leading to the increase in reaction times. However, this also indicates that difficulties with L2 gender may not be due to the need to inhibit the L1 gender during L2 production per se but rather due to difficulties learning and producing the L2 gender when it is different from L1 gender. Gender is a difficult grammatical structure to acquire, so the interaction that many studies observe could simply be due to difficulties L2 gender.

However, so many studies have found evidence of L1 influence in L2 gender production, that there is a great deal of evidence supporting L1 influence. Kupisch, Akpinar, and Stöhr (2013) also found support of L1 influence. They tested German-French bilinguals, whose languages contain gender systems that are unequal in the number of different gender categories, as well as the distribution of gender assignment. French has two genders, masculine and feminine (e.g. *le chien*$_M$ 'the dog' vs. *la porte*$_F$ 'the door'), which is marked on determiners, adjectives (which appear after the noun except in certain cases), some quantifiers, and past participles. Gender is not

evenly distributed; 61% of words are masculine and 39% of words are feminine. On the other hand, German has three genders – masculine, feminine and neuter (e.g. *der Hund*$_M$ 'the dog' vs. *die Brücke*$_F$ 'the bridge' vs. *das Pferd*$_N$ 'the horse'). Determiners (in combination with number and case information), relative and question pronouns, and attributive adjectives are marked for gender in German. As in French, the distribution of gender assignment is not evenly distributed, as 50% of words are masculine, 30% are feminine, and 20% are neuter (Kupisch et al., 2013).

The German-French participants in Kupisch et al. (2013) varied based on their number of years of L2 immersion and age of acquisition. Participants in the study completed an acceptability judgment task and an elicited production task. Results suggested that the age of acquisition does not affect how well the participants were able to use gender in French, but immersion (whether the participant learned French living in a German-speaking or a French-speaking society) slightly influenced performance in terms of native-like accuracy in gender agreement. This experiment, however, only focused on masculine and feminine nouns, neglecting nouns that are neuter in German. Further, their study did not include any online measures of language production.

Current Study

The current study builds upon Kupisch et al. (2013) using methodologies from previous studies (e.g., Costa et al., 2003; Paolieri et al., 2010; Bordag and Pechmann, 2003). Despite the large number of studies that have looked at the processing of L2 grammatical gender, open questions remain. For example, of the studies mentioned above (Bordag & Pechmann, 2003; Paolieri et al., 2010) and others (e.g., Salamoura & Williams, 2007), the cross-linguistic comparisons have involved gender systems that have the same number of grammatical gender categories. Others look at gender systems with different numbers of grammatical gender categories, such as the study by Lemhöfer et al. (2008) testing German-Dutch participants, but

where there is a high level of overlap in gender assignment across the two languages. Although Kupisch et al. (2013) observed the same population as the current study, their study focused on how L2 immersion and age of acquisition influence the acquisition of the L2 gender system, whereas the present study investigates how differences between French and German grammatical gender systems influence L2 processing. In contrast to Kupisch et al. (2013), the present study focuses specifically on words that are neuter in German, referred to here as the *No Gender Match* condition (e.g. German: *das Pferd*$_N$ vs. French: *le cheval*$_M$ meaning 'the horse'; or German: *das Lineal*$_N$ vs. French: *la règle*$_F$ meaning 'the ruler'). Investigating the production of L2 French gender among German L2 learners of French, the present study aims to uncover further evidence that the L1 influences L2 gender learning, while additionally observing how German-French L2 speakers produce noun phrases in L2 French that have no direct corresponding gender in their L1 German (i.e., the *No Gender Match* condition).

As stated previously, German and French have different grammatical gender systems. German has a three-gender system - masculine, feminine, and neuter - while French has a two-gender system - just masculine and feminine. Because of these differences, this study will investigate whether there is an influence of the L1 gender system during the production of L2 gender, particularly focusing on words that are neuter in German. The experiment will do so through a picture-naming study that measures the accuracy and reaction times of advanced German- L2 French learners based on stimuli in three different categories. These categories include congruent items (gender is the same in both languages), incongruent items (gender is different between the two languages, e.g. if gender is masculine in German, it is feminine in French and vice versa), and No Gender Match items (gender in German is neuter and either masculine or feminine in French).

Based on the findings of previous studies, such as Bordag and Pechman (2007) and Paolieri et al. (2010), which report that L1 gender assignment influences L2 gender production,

we predict to see higher error rates and longer RTs on words in the incongruent condition, and that participants will be more accurate with faster RTs on words in the congruent condition. The no gender match condition could show one of two different patterns. The first of these would be lower accuracy rates and longer RTs, similar to the incongruent condition, since the items in the no gender match condition have two conflicting genders between German and French. This would suggest that participants experience difficulty with items in the No Gender Match condition because genders are not congruent across languages. Alternatively, items in the no gender match condition could pattern like neither the congruent condition nor the incongruent condition, meaning that the accuracy and RTs in the no gender match condition would lie somewhere between the congruent and incongruent conditions. More specifically, participants would not be as accurate or as fast to produce items in the no gender match condition as they are in the congruent condition, but they will also not be as inaccurate or slow as they are in the incongruent condition. This would suggest that nouns that are neuter gender in the participants' L1 German would neither directly inhibit nor facilitate the production of masculine or feminine nouns in the L2 French.

Methods

Participants

Thirty-four German-French speaking participants were recruited in Berlin, Germany. Twenty of these participants were found via a listserv at Humboldt Universität that is sent to the entire student population. Four were recruited on Craigslist, and ten were found by word-of-mouth, either by the experimenter visiting upper-level French classes at the university or finding acquaintances of colleagues.

Fifteen participants were excluded due to low French proficiency and poor performance on the task (identified less than 5 items correctly in each condition). Two participants were excluded because German was not their first language. Thus, the bare-noun task and the determiner-noun phrase task results reported here are based on the results from 17 participants (bare-noun task: average age: 25.8; age range 19-35; 11 females, 6 males – determiner-noun phrase task: average age: 25.3; age range 19-35; 13 females, 4 males).

Materials

The stimuli consisted of 60 black and white images taken from the International Picture Naming Project database (Szekely et al., 2004), or other identifiable clipart pictures taken from the internet. All were high frequency, concrete nouns, and none had natural gender (e.g., no humans). The stimuli varied according to gender congruency between German and French. Twenty items had the same gender in German and French, as in (1). Twenty items had conflicting

gender between German and French, as in (2), meaning that if gender was masculine in German, it was feminine in French and vice versa. Twenty items were neuter gender in German, as in (3), meaning that there was no corresponding gender in French.

Figure 1: Example Stimuli

(1) **Congruent Gender:**
English: the *flower*
French: *la.*$_F$ *fleur*
German: *die.*$_F$ *Blume*

(2) **Incongruent Gender:**
English: the *bridge*
French: *le.*$_M$ *pont*
German: *die.*$_F$ *Brücke*

(3) **No Gender Match:**
English: the *heart*
French: *le.*$_M$ *coeur*
German: *das.*$_N$ *Herz*

Furthermore, within each condition there were an equal number of masculine and feminine stimuli in both languages. The only condition in which this could not be done was the No Gender Match condition, which had 8 feminine nouns and 12 masculine nouns in French. This was due to the limited number of high frequency, concrete nouns that are feminine in French and neuter in German. Stimuli were carefully balanced so that the 60 final stimuli included no cognates between French and German and no French words that started with vowels, as words that begin with vowels are all preceded by the same determiner, *l'*, in French. Across conditions, there was no significant difference in word frequency in French or German, or the number of syllables in French or German (all *p*s > .09). There was no difference in naming agreement for the target pictures in German ($p = .229$) but there was a significant difference in naming agreement in

French ($F(2) = 4.53$; $p = .017$), because naming agreement was lower for the Incongruent and No Gender Match conditions compared to the Congruent condition.

Procedure

The experiment lasted about 90 minutes and included six picture-naming tasks that were split into two blocks. Participants were tested individually while the experimenter remained in the room. Participants' voices were recorded by a voice recorder, and their reaction times were recorded using E-Prime 2.0 (Schneider, Eschman, & Zuccolotto, 2002).

The first block was comprised of three different tasks that the participants completed in French. These tasks included a bare-noun task, a determiner-noun task in which participants named the determiner and the noun, and an adjectival-phrase task, in which participants named the determiner plus an adjective plus the noun. All three tasks included the same 60 stimuli. For each task and for each participant, the stimuli were presented in a different randomized order. However, each participant completed the three tasks in the same order: first the bare-noun task, then the determiner-noun task, then the adjectival phrase task.

Before each task, the participant received vocal and written instructions in French. They completed a short practice set before beginning the actual task. Before every trial a black fixation cross appeared on the screen for 1000 ms. Then the target picture appeared for a maximum of 4000 ms. Participants were instructed to name the picture as quickly and accurately as possible. Once the voice key picked up the participant's response, the image disappeared and the fixation cross appeared again for 1000 ms before displaying the next image.

After completing these three tasks, participants completed the Boston Naming Task (BNT) in French and the O-Span task, also in French, as a measure of working memory (adapted

from Turner & Engle, 1989). The O-Span only required words to be memorized, but all directions were in French.

The final block of tasks replicated the first three picture-naming tasks exactly. They included the same images and the same instructions. However, participants were instructed to complete the bare-noun task, the determiner-noun task, and the adjectival phrase task in their L1 German.

The bare-noun task was used as a baseline to determine participant L2 French proficiency and to expose participants to the target stimuli pictures before completing the determiner-noun task. As the instructions stated, participants identified each object that appeared on the screen using only the French noun, without any accompanying article.

The noun-phrase task was comprised of the same stimuli as the bare-noun task. However, the images appeared in a different, randomized order. For this task, participants were instructed to state the appropriate determiner, *le* or *la*, followed by the noun. Although participants were told to use the definite determiners, *le* and *la,* responses using the indefinite determiners *un* and *une* were included in the accuracy and RT analyses.

The final task was the adjectival-phrase task. The same 60 stimuli from the previous two tasks were used. However, 50% of the stimuli were enlarged and 50% were shrunk so that the size difference was noticeable. They were counterbalanced so that half of the participants saw, for example, a small flower, while the other half saw a large flower. Participants had to identify each item first stating the definite determiner, *le* or *la*, followed by an adjective describing the size of the image, and finally followed by the noun. Thus, an expected response for a large flower would be *la grande fleur,* "the big flower". This task, however, was excluded from data analysis due to low accuracy rates; therefore, it will not be discussed further in this thesis.

Results

Table 1: Bare-Noun Task *n*=17 - Accuracy and Reaction Times by Condition

	Accuracy (% Correct)		Reaction Times (in ms)	
Condition	*Mean*	*SD*	*Mean*	*SD*
Congruent	66.9	11.3	1483	320
Incongruent	58.1	13.9	1650	402
No Gender Match	66.3	14.9	1579	339

A total of 324 data points (35.96 % of data) were excluded from the final accuracy analyses. Data was excluded due to voice key errors or lack of a response or incorrect identification of a noun. Responses with an article before the noun were included as long as the participant said the correct noun. Seven items had to be excluded due to less than 30% accuracy naming. After excluding these 7 items, 19 Congruent items, 16 Incongruent items, ad 18 No Gender Match items remained as part of the analyses. Data points were based on the responses of 17 participants who accurately and timely identified at least five items in each condition on the bare-noun task. [1]

The accuracy data were analyzed using a repeated-measures ANOVA with Condition (Congruent, Incongruent, No Gender Match) as a within-participants variable.[2] Results revealed a marginally significant difference across conditions, $F(1.5, 24) = 2.84$, $p = .091$). As seen in Table 1, descriptively, participants were more accurate in the Congruent and the No Gender Match conditions than the Incongruent condition.

[1] The same items were used in all analyses for both the bare-noun and the determiner-noun tasks

[2] For all ANOVAs if sphericity was violated, the Greenhouse-Geiser correction was applied to the degrees of freedom and the corrected *p*-value is reported

418 data points (45.84 %) were excluded from the final RT analyses. Data points were excluded due to voice key errors, lack of a response, incorrect identification of an item, and the addition of an article before the noun. Data was coded in a numerical system where lack of a response and incorrect identification of a noun were marked as 0. Items with a gender added before the noun were coded as 5. After initially excluding all data marked with these numbers, we ran further trims excluding all data with RTs <300 or >3000. Then we found subject outliers, defined as greater than 2.5 standard deviations above or below the mean RT for that participant, and also excluded items outside of these boundaries.

A repeated-measures ANOVA treating Condition as a within-participants variable revealed a marginally significant difference in reaction times across conditions, $F(2, 32) = 3.06$, $p = .061$. As seen in Table 1, the No Gender Match condition elicited RTs that were between the other two conditions and, once again, the longest RTs were in the Incongruent condition.

Table 2: Bare-Noun Task *n*=13 - Accuracy and Reaction Times by Condition

	Accuracy (% Correct)		Reaction Times (in ms)	
Condition	*Mean*	*SD*	*Mean*	*SD*
Congruent	67.2	11.2	1447	345
Incongruent	58.2	13.8	1657	457
No Gender Match	68.0	14.4	1506	311

A second set of analyses were performed in the exact same manner on only the participants who named at least five items correctly in each condition on both the bare-noun task and the determiner-noun task. For these ANOVAs, 13 participants were included. A total of 243 data points (35.27% of data) were excluded from the final accuracy analyses. Exclusions occurred due to voice key errors, lack of a response or incorrect identification of a noun. A repeated-measures ANOVA showed no significant difference across conditions for naming accuracy, $F(2, 24) = 2.175$, $p = .136$.

316 data points (45.86% of data) were excluded from the final RT analyses. Exclusions were based on voice key errors, lack of a response, incorrect identification of a noun, and the addition of an article before the noun. A repeated-measures ANOVA revealed a significant difference across conditions, $F(2, 24) = 3.69$, $p = .040$. To explore this main effect, we ran follow-up t-tests comparing each condition to one another. The t-tests revealed a significant difference between the Incongruent and Congruent conditions, $t(12) = 2.85$, $p = .015$). But there was no significant difference between the Incongruent and No Gender Match conditions, $t(12) = 1.59$, $p = .139$), nor between the Congruent and No Gender Match conditions, $t(12) = 0.88$, $p = .394$. As indicated in Table 2, the Congruent condition elicited the lowest RT latencies ($M = 1447$ ms), while the Incongruent condition elicited the highest RT latencies ($M = 1657$ ms). The No Gender Match RTs lie between the other two conditions ($M = 1506$ ms). These results, by paralleling the findings of the n=17 bare-noun analyses, strengthen them.

Table 3: Determiner-Noun Phrase *n*=17 Gender Assignment Errors (%)

Condition	Correct Noun /Correct Gender %	Correct Noun/Wrong Gender %	Incorrect Noun %
Congruent	219(67.8%)	17(5.3%)	87(26.9%)
Incongruent	154(57.7%)	27(9.9%)	88(32.4%)
No Gender Match	222(72.5%)	7(2.3%)	77(25.2%)

Table 4: Determiner-Noun Phrase Task (*n*=17) – Gender Assignment Accuracy and Reaction Times by Condition

	Gender Assignment Accuracy (% Correct)		Reaction Times (in ms)	
Condition	*Mean*	*SD*	*Mean*	*SD*
Congruent	92.3	8.2	1099	191
Incongruent	85.1	10.9	1212	239
No Gender Match	96.2	6.7	1149	211

In the determiner-noun phrase task a total of 276 data points (28.93% of data) were excluded from the final accuracy analyses. As with the bare-noun task, excluded items included voice key errors, incorrectly identified nouns, and incorrectly assigned gender. Data points were based on the responses of 17 participants who accurately and timely identified at least five items in each condition on the determiner-noun phrase task.[3]

A repeated-measures ANOVA treating Condition as a within-participants variable revealed a significant difference between conditions ($F(2, 32) = 8.152$, $p = .001$). To explore this main effect, we ran follow-up t-tests comparing each condition to one another. The t-tests revealed a significant difference between the Congruent and Incongruent conditions, and the No Gender Match and Incongruent conditions [Congruent versus Incongruent: $t(16) = 2.135$, $p = .049$; No Gender Match versus Incongruent: $t(16) = 4.124$, $p = .001$]. The difference between the Congruent and the No Gender Match conditions approached significance ($t(16) = 1.796$, $p = .091$).

394 data points (41.3%) were excluded from the final RT analyses. Exclusions were based on voice key errors, lack of a response, incorrect identification of a noun, and incorrect assignment of gender. Data was coded in a numerical system where lack of a response and incorrect identification of a noun were marked as 0. Gender errors were marked as 2. After initially excluding all data marked with these numbers, we ran further trims excluding all data with RTs <300 or >3000. Then we found subject outliers defined as greater than 2.5 standard deviations above or below the mean RT for that participant, and also excluded items outside of these boundaries.

Turning to the RT analyses, a repeated-measures ANOVA treating Condition as a within-participants variable revealed a main effect of Condition ($F(2,32) = 9.33$, $p = .001$). To explore

[3] The 17 participants included in the determiner-noun task analyses were not the same participants as those included in the bare-noun task.

this main effect, we ran follow-up t-tests comparing each condition to one another. The *t*-tests revealed a significant difference between the Congruent and Incongruent conditions, and the No Gender Match and Incongruent conditions [Congruent versus Incongruent: $t(16) = 4.19$, $p = .001$; No Gender Match versus Incongruent: $t(16) = 2.49$, $p = .001$]. The difference between the Congruent and the No Gender Match conditions was approached significance ($t(16) = 1.9$, $p = .075$).

Table 5: Determiner-Noun Phrase *n*=13 - Gender Assignment Errors (%)

Condition	Correct Noun/Correct Gender %	Correct Noun/Wrong Gender %	Incorrect Noun%
Congruent	169(68.4%)	12(4.9%)	66(26.7%)
Incongruent	123(59.1%)	18(8.7%)	67(32.2%)
No Gender Match	174(74.4%)	5(2.1%)	55(23.5%)

Table 6: Determiner-Noun Phrase Task *n*=13 – Gender Assignment Accuracy and Reaction Times by Condition

	Gender Assignment Accuracy (% Correct)		Reaction Times (in ms)	
Condition	*Mean*	*SD*	*Mean*	*SD*
Congruent	92.8	8.7	1070	184
Incongruent	87.3	10.4	1165	209
No Gender Match	96.6	5.8	1117	201

Similar to the bare-noun task, we also conducted a second analysis on the determiner-noun phrase task with the 13 participants that named at lease five items in each condition correctly on both the determiner-noun phrase and the bare-noun tasks. A total of 188 data points (27.29% of data) were excluded from the final accuracy analyses. Results were very similar to those of the determiner-noun phrase analyses including all 17 participants, strengthening those findings.

A repeated-measures ANOVA revealed a significant difference in accuracy between the three conditions ($F(2, 24) = 4.25$, $p = .026$). To explore this main effect, we ran follow-up t-tests

comparing each condition to one another. The t-tests revealed a significant difference between the No Gender Match and Incongruent conditions, ($t(12) = 3.24$, $p = .007$). The differences between the Congruent and Incongruent and the Congruent and No Gender Match conditions were not significant [Congruent versus Incongruent: $t(12) = 1.4$, $p = .187$; Congruent versus No Gender Match: $t(12) = 3.24$, $p = .182$].

From the final RT analyses 262 data points (38.03%) were excluded. These exclusions were based on voice key errors, lack of a response, incorrect identification of a noun, incorrect assignment of gender, and outlier RTs.

Results from a repeated-measures ANOVA reveal a main effect of condition ($F(2, 24) = 4.25$, $p = .026$). To explore this main effect, we ran follow-up t-tests comparing each condition to one another. The t-tests revealed a significant difference between the Congruent and Incongruent conditions, ($t(12) = 3.25$, $p = .007$). The differences between the Congruent and Incongruent and the Congruent and No Gender Match conditions were not significant, although the Incongruent and No Gender Match *p*-value approached significance [No Gender Match versus Incongruent: $t(12) = 1.9$, $p = .081$; Congruent versus No Gender Match: $t(12) = 1.56$, $p = .144$].

We also performed a post-hoc analysis to observe whether there may be gender bias in the No Gender Match condition. Often, L2 learners favor one gender over the other (e.g., they choose a masculine gender label more frequently than a feminine gender label) if they are unsure of the correct gender. However, the naming latencies for each condition strongly suggested that there was no gender bias, so we saw no need to investigate this possibility further.

General Discussion

This experiment yielded some interesting findings concerning the interaction of L1 and L2 gender systems during L2 production, notably that words that have no equivalent gender assignment between the L1 and the L2 are produced differently than words that have congruent or incongruent grammatical gender across languages. Significant differences in the determiner-noun phrase task indicated that participants were less accurate and slower to name items in the Incongruent condition while they were fastest to name items in the Congruent condition. The faster RTs in the Congruent condition suggests that the L1 gender facilitates L2 gender production when gender is the same across languages, while the longer RTs in the Incongruent condition indicate that the L1 gender inhibits L2 gender production when the gender conflicts across languages. These results parallel the findings of previous L2 studies, such as Paolieri et al. (2010) and Bordag and Pechman (2007).

The RTs in the No Gender Match condition, however, lie between those of the Congruent and Incongruent genders, suggesting that there is no direct interaction between German neuter gender and French masculine or feminine genders. RTs show that there is neither facilitation nor inhibition from the L1 to the L2 in this condition. If L1 gender influences L2 gender production equally across the board, and this influence manifests itself in difficulties learning the L2 gender of words that do not have the same gender in both languages, we would have expected to see results similar to those seen in the Incongruent condition. However, the different results suggest that this gender condition is processed differently from the Congruent or Incongruent conditions.

These findings suggest that there is L1 influence on L2 gender production, and it is not solely due to difficulties learning the language. If the No Gender Match condition had behaved like the Incongruent condition, it would have suggested that participants have difficulty learning and producing all gender that is incongruent across languages. This, however is not the case. Further, descriptively speaking, participants were most accurate in the No Gender Match condition – a finding that at present has no clear explanation—but one that warrants more research.

The results of the bare-noun task analyses are surprising because they show the same pattern of results as the determiner-noun phrase task. Although not all conditions showed statistically significant differences, the trend still emerged that the Congruent condition elicited the fastest RTs while the Incongruent condition elicited the slowest, and the RTs for the No Gender Match condition were between both. Because participants are not overtly producing a gender-marked determiner in the bare-noun task, one might expect that the gender of the word would not influence the accuracy rates and RTs when naming the items. However, since the bare-noun data shows the same trends as the determiner-noun task, this may indicate that grammatical gender still influenced production in the bare-noun task. Therefore, it is possible that although participants are not producing the gender-marked articles, gender is still active as they retrieve target words in their L2; this, as a result, influences their accuracy and their reaction times.

We must, nonetheless, also acknowledge additional factors that may play a role in these results. The relatively low accuracy rates in this study suggest that participants were not near-native speakers of French, and therefore should still be considered L2 learners. As a result, it is not possible to completely disentangle the role of L1 interaction from the difficulty in learning the incongruent L2 gender or the inhibition of L1 gender during production. The low number of participants may also affect the outcomes of this study due to low statistical power. Since a large number of participants were excluded due to low proficiency, this study was left with fewer participants. Lastly, although stimuli were carefully balanced, more items had to be deleted from

the Incongruent condition than the Congruent and No Gender Match conditions. It is possible that this may have also affected the results.

To address these limitations, we intend to complete some follow-up studies to strengthen the results of the present study, and have made arrangements to collect data from additional German-French participants in Germany. In addition, a parallel study to this one will investigate how English L2 learners of French perform on the same experimental tasks. Because the English-French learners do not have gender in their L1, they are expected to perform with equal accuracy and RTs across all conditions. If this is the case, their results may shed further light on the nuances of the No Gender Match condition. If the accuracy and RTs of this study are similar to those of the German-French learners in the No Gender Match condition, they many further strengthen the evidence that the L1 influence we see in this study is not due to difficulty learning conflicting gender in the L2. Since these results highlight the difficulty L2 learners have with conflicting gender in the L2, they may play a role in classroom learning. A teacher may be more aware of the difficulty students may have with these items and focus on them more in class. They may even change standard curriculum to facilitate learning for items that conflict across languages.

Additionally, the results from this study contribute to current debates in the research literature on bilingual gender, which has not yet studied items that have no direct equivalent in the L2. It may spur further research in languages with this these types of nouns to explore other ways the L1 and L2 could interact in this when there is no direct L2 equivalent. There has been, however a recent study by Rachel Klassen, that has not yet been published, observed Spanish-German L2 learners and found very similar results as the current study. Nonetheless, further research is needed to distinguish exactly how the brain treats these types of lexical items. A more proficient population may also be tested to further tease apart the question as to whether the L1

influence observed in this study is due to difficulties learning conflicting gender in the L2 or inhibition of the L1 gender of these lexical items.

Appendix A

List of Stimuli

Table 7: Incongruent Stimuli

French	German	English
Balai	*Besen*	Broom
Bougie	*Kerze*	Candle
Bouteille	*Flasche*	Bottle
Cerise	*Kirsche*	Cherry
Chèvre	*Ziege*	Goat
Chien	*Hund*	Dog
Cloche	*Glocke*	Bell
Coquille	*Muschel*	Shell
Fleur	*Blume*	Flower
Fraise	*Erdbeere*	Strawberry
Fromage	*Käse*	Cheese
Gateau	*Kuchen*	Cake
Papillon	*Schmetterling*	Butterfly
Parapluie	*Regenschirm*	Umbrella
Pied	*Fuss*	Foot
Porte	*Tür*	Door
Pouce	*Daumen*	Thumb
Rue	*Straße*	Road
Singe	*Affe*	Monkey
Stylo	*Kuli*	Pen

Table 8: Incongruent Items

French	German	English
Canard	*Ente*	Duck
Ceinture	*Gürtel*	Belt
Clé	*Schlüssel*	Key
Collier	*Halskette*	Necklace
Cuillère	*Löffel*	Spoon
Flèche	*Pfeil*	Arrow
Fouet	*Peitsche*	Whip
Fumée	*Rauch*	Smoke
Fusil	*Gewehr*	Rifle
Lune	*Mond*	Moon
Monde	*Welt*	World
Nuage	*Wolke*	Cloud
Phare	*Leuchttum*	Lighthouse
Pluie	*Regen*	Rain
Sifflet	*Pfeife*	Whistle
Pont	*Brücke*	Bridge
Tableau	*Tafel*	Chalkboard
Tambour	*Trommel*	Drum
Tondeuse	*Rasenmauer*	Lawnmower
Valise	*Handkoffer*	Suitcase

Table 9: No Gender Match Items

French	German	English
Carré	*Quadrat*	Square
Cerf	*Reh*	Deer
Cerveau	*Gehirn*	Brain
Château	*Schloß*	Castle
Chemise	*Hemd*	Shirt
Cheval	*Pferd*	Horse
Cochon	*Schwein*	Pig
Coeur	*Herz*	Heart
Couteau	*Messer*	Knife
Fenêtre	*Fenster*	Window
Feuille	*Blatt*	Leaf
Jambe	*Bein*	Leg
Lapin	*Kaninchen*	Rabbit
Moufette	*Stinktier*	Skunk
Mouton	*Schaf*	Sheep
Poule	*Hühn*	Chicken
Règle	*Lineal*	Ruler
Robe	*Kleid*	Dress
Roue	*Rad*	Wheel
Toit	*Dach*	Roof

BIBLIOGRAPHY

Bordag, D., & Pechmann, T. (2007). Factors influencing L2 gender processing. *Bilingualism: Language and Cognition, 10(3),* 299-314.

Costa, A., Kovačic , D., Franck, J. & Caramazza, A. (2003). On the autonomy of the grammatical gender systems of the two languages of a bilingual. *Bilingualism: Language and Cognition, 6(3),* 181–200.

De Bot, K. (1992). A bilingual production model: Levelt's speaking model adapted. *Applied Linguistics*, *13*, 1–24.

Kroll, J. F. & Tokowicz, N. (2005). Models of bilingual representation and processing: Looking back and to the future. *Handbook of bilingualism: Psycholinguistic approaches,* 531–553. New York: Oxford University Press.

Kupisch, T., D. Akpinar & A. Stöhr. 2013. Gender assignment and gender agreement in adult bilingual and second language speakers of French. *Linguistic Approaches to Bilingualism 3(2),* 150-179.

Lemhöfer, K., Spalek, K., & Schriefers, H. (2008). Cross-language effects of grammatical gender in bilingual word recognition and production. *Journal of Memory and Language, 59(3),* 312-330.

Paolieri,D., Cubelli, R., Macizo, P., Bajo M. T., Lotto, L., & Job, R. (2010). Grammatical gender processing in Italian and Spanish Bilinguals. *The Quarterly Journal of Experimental Psychology, 63(8),* 1631-1645.

Sabourin, L.L., Stowe, L.A., & de Haan, G.J. (2006). Transfer effects in learning and L2 grammatical gender system. *Second Language Research, 27(1)*, 1-29.

Salamoura, A., and Williams, J. N. (2007). The representation of grammatical gender in the bilingual lexicon: Evidence from Greek and German. Bilingualism: *Language and Cognition 10,* 257-275.

Schneider, W.; Eschman, A.; Zuccolotto, A. (2002). E-Prime User's Guide. Pittsburgh: Psychology Software Tools, Inc.

Szekely et al., 2004, *Journal of Memory and Language*, *51(2),* 247-250.

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

ACADEMIC VITA

Leah Pappas
109 Pineridge Rd
Pittsburgh, PA 15237
leah.j.pappas@gmail.com

EDUCATION

The Pennsylvania State University **2010-Present**
Major: French and Francophone Studies
Major: Linguistics
Minor: Spanish
Expected Graduation Date: May 2014

HONORS

Student Marshal, The Department of French and Francophone Studies — Spring 2014
NSF-PIRE Fellowship Recipient — December 2012-Present
Paterno Fellows Student — Fall 2010-Present
Schreyer's Honor's Scholar — Spring 2012-Present
President's Freshman Award — Fall 2010
Dean's List — Fall 2010-Present

RESEARCH EXPERIENCE

Bilingualism and Language Development Lab **Fall 2012- Present**
Center for Language Science, The Pennsylvania State University
Advisor: Dr. Janet Van Hell

- Code data from various tasks using the E-prime suite, Microsoft Excel, and R
- Attend weekly lab meetings in which relevant research in the field is discussed
- Independently run ERP and Behavioral experiments

Center for German Language and Linguistics **May 2013 – July 2013**
Humboldt, Universität, Berlin, Germany
Advisor: Dr. Katharina Spalek

- Run my own study with German-French learners
- Attend weekly lab meetings in which relevant research in the field is discussed
- Code and analyze data from the study

Language and Cognition Lab **June 2012 - August 2012**
Center for Language Science, The Pennsylvania State University
Advisor: Dr. Judith Kroll

- Code data from various tasks using the E-prime suite, Microsoft Excel, and R

- Independently run Behavioral and ERP experiments
- Advertise and schedule studies with participants

Second Language Acquisition Lab **January 2011- Present**
Center for Language Science, The Pennsylvania State University
Advisor: Dr. Carrie Jackson

- Code data from various tasks using the E-prime suite, Microsoft Excel, and R
- Attend lab meetings in which relevant research in the field is discussed
- Independently run Behavioral experiments

WORK EXPERIENCE

Teaching Assistant **January 2014-May 2014**
The Pennsylvania State University

- Prepare in-class activities
- Prepare and administer daily attendance quizzes
- Serve as a mentor and a second resource to students who have questions about the material

French Tutor **September 2012-Present**
The Pennsylvania State University

- Tutor lower level French students
- Answer questions, review lessons, practice speaking

Direct Care at The ARC of Centre County **June 2012-December 2013**

- Direct care with people with mental disabilities, helping them with daily activities
- Paperwork and documentation as well as communication with higher-ups

Internship: Teaching Assistant **January 2012-March 2012**
L'Ecole Saint-Odile, Montpellier France

- Worked with a French elementary school teacher of English
- Experienced the paperwork part of teaching, put together lessons, games and helped grade papers
- Independently taught English to four classes a day with children from the ages of 4-11
- Communicated with parents and other teachers in the school

ACTIVITIES

Study Abroad: Montpellier, France **Spring 2012**

- Attended Montpellier III, Université Paul Valéry, taking classes with French students and professors
- Lived with a French family in order to practice speaking and learn the culture
- Involved in sports at Paul Valéry, immersed in various exercise classes
- Attended weekly conversation gatherings, organized meetings where people from different countries practice speaking another language with a partner.

Director/Actor with No Refund Theater **August 2012-Present**

- Secretary September 2013-Present: send out weekly emails and updates
- Perform in about 2 shows each semester
- Directed and produced *Arsenic and Old Lace* by David Kesselring

THON Family Relations Chair **September 2013-Present**

- In September 2013, created an organization that raises money to help treat children with cancer
- Organize and conduct fundraisers
- Maintain contact with and provide emotional support to a family paired to our organization with a child suffering from cancer

Thespians

- Worked with others to put on charity performances for children and people suffering from AIDS