LES FAUX-AMIS: INVESTIGATING LEXICO-SEMANTIC AMBIGUITY ACROSS TWO LANGUAGES[*]

Michèle Burkholder
University of Ottawa

1. Introduction

When acquiring a second language (L2), a significant task is learning a new set of lexical items, a process that involves making thousands of new correspondences between form and meaning. This task, however, can seem somewhat less daunting when the lexicon of the language being acquired overlaps to some degree with that of the learner's first language (L1). For example, learning the French word *table* is very straightforward for L1 speakers of English, as it is the translation equivalent of the English word *table*.

Words that overlap in form across two languages, however, do not always have complete semantic overlap. Borrowed words can often be imported with slightly different senses and, because languages are in constant evolution, the meaning of borrowings and also of cognates (words with a common genetic origin) can diverge over time. Further, some words that share their form across two languages do so by sheer coincidence, e.g., the French word *coin* means "corner", not "coin" in English. Together, these words are referred to as "False Friends", as they are notoriously deceiving to the L2 learner.

Previous psycholinguistic studies have shown that, within a language, all meanings of homonyms are initially accessed from memory (e.g. Swinney 1979). Interestingly, the same has been found across a bilingual's two languages; words with the same form in both languages can also activate each other in memory (e.g. Dijkstra et al. 2000). For false friends, this would lead to the activation of semantically inappropriate words.

The aim of this study is thus to examine how word meanings of false friends in the second language (L2) interfere in L1 word processing, and how this changes as a function of L2 proficiency. English-French bilinguals and English monolinguals were tested using a speeded semantic relatedness task, where they rated the similarity of false friend word meanings in their L1. Before presenting this study, however, the different types of false friends will be defined, and issues pertaining to their processing will be discussed.

2. Background

2.1 Defining the terms: Words that overlap in form and meaning

English and French have a substantial number of lexical items that overlap in form[1], due in part to their common Indo-European origin, and also to the fact that they have been in significant contact over the course of history. Lexical items that overlap in form, however, can vary in terms of the amount of semantic overlap that they share, and this can be envisaged to exist along a continuum. At one extreme are cognates, which have

[*] I would like to thank my supervisor, Laura Sabourin, the members of the ERP Linguistics Lab, especially Santa Vīnerte, Myriam Lapierre, Fauve Duquette-Laplante, and Natalie Ho.

[1] The focus here will be on shared orthographic form, as the experimental task uses the visual modality.

Actes du congrès annuel de l'Association canadienne de linguistique 2015.
Proceedings of the 2015 annual conference of the Canadian Linguistic Association.

complete semantic overlap. Note that the term "cognate" will be defined here as has traditionally been done in psycholinguistics: translation equivalents which share a substantial amount of form across the two languages (Sunderman and Schwartz 2008). For example, the French translation of the English word *table* /tej.bl/ is *table* /tabl/.

The rest of the semantic overlap continuum is populated by false friends, and these can be classified into different types, four of which were investigated in this study. The first type are "Interlingual Homographs" (ILHs), which are found at the other extreme of the continuum, opposite to cognates. Despite their significant overlap in form, the referents of ILHs have completely unrelated meanings in their respective languages; the common form shared between such words is purely accidental (at least synchronically). For example, while the orthographic forms of the English word *coin* and the French word *coin* ("corner") overlap completely, there is no common element to their meanings.

Like interlingual homographs, "False Cognates" (FCs) have different referents in each language, however, crucially, these referents are semantically related. For example, the English word *library* refers to a place where books are borrowed (in French this is a *bibliothèque*), whereas the French word *librairie* refers to a place where books are bought (in English a *bookstore*). These are two distinct concepts that are represented in each language by two distinct words that are not used interchangeably.

False friends can also be categorized with respect to how many of their meanings are shared between the two languages. Many false friends have only one core meaning in each language, e.g. *library/librairie* and *coin/coin*. Such cases will be referred to as *Full* False Friends. However, as homonymy and polysemy are common within a language, sometimes the multiple meanings associated to a single word-form in one language are not all translated into the other language using a single word-form. Such cases will be referred to as *Partial* False Friends: word pairs that have one referent common to both languages, and an additional referent in only one of the languages.

In the case of Partial ILHs, this second referent has very little in common semantically with the shared meaning. For example, the English word *peach* refers only to the stone fruit, whereas the French word *pêche* means both "peach" and "fishing". In contrast, for Partial FCs, this second referent has a significant amount of common semantic properties with the shared referent. For example, the English word *corpse* refers only to a deceased body, yet the French word *corps* means both "corpse" and "body".

These four types of false friends (Full ILHs, Partial ILHs, Full FCs, and Partial FCs) will be the objects of investigation in the study presented here. While previous studies have examined how cognates and ILHs are stored and processed by bilinguals, few have looked into these more complex types of false friends. Yet these are the kinds of false friends that are potentially the most problematic for bilinguals and L2 learners.

2.2 Bilingual lexical processing

2.2.1 The bilingual mental lexicon and L2 proficiency effects

Speakers who know two languages must have two sets of lexical items; however, are these two lexicons represented separately in memory, or in a single integrated system? This issue has been a primary area of psycholinguistic investigation for decades (e.g. Potter et al. 1984; Kroll and Stewart 1994; Dijkstra and van Heuven 1998, 2002; Sabourin et al. 2014). The emerging consensus is that at least highly proficiency bilinguals who learned both languages at an early age do indeed have a single memory

system in which words from both languages are stored. As such, most of the prominent models of the bilingual lexicon assume overlap between the two lexical stores at least at the semantic level of representation (e.g. Kroll and Stewart 1994, Finkbeiner et al. 2004), if not at all levels of representation, including orthographic and phonological levels, for example the Bilingual Interactive Activation (BIA) models of Dijkstra and van Heuven (1998, 2002). It is this latter type of model that would best account for the cross-language lexical interference effects expected in the present study, as it is the common orthographic/phonological form of false friends that would trigger the cross-language activation that is the source of the interference effects.

An aspect of the bilingual lexicon that is still under investigation is the extent to which the system is shared for bilinguals who acquired their L2 at different ages and who have different levels L2 proficiency, amongst other factors (e.g. Dimitropoulou et al. 2011, Sabourin et al. 2014). Crucially related to this issue is an asymmetry whereby the L1 has a much stronger impact on L2 processing than vice versa (Duñabeitia et al. 2010). As such, investigating L2 effects on L1 processing can shed the most light into the issues of how factors such as L2 proficiency impact bilingual lexical processing.

The current study thus examines interference effects from the L2 to the L1 in order to investigate the role that L2 proficiency plays in the processing of false friends. For example, it is possible that higher L2 proficiency leads to more interference because L2 word meanings have stronger mental representations and so have a greater potential impact on the L1 (e.g. Jiang 1999). On the other hand, it may be the case that higher L2 proficiency will result in *less* interference, as highly proficient bilinguals are better at controlling activation of their two languages and can therefore limit the amount of cross-language interference from both the L1 and the L2 (e.g. Abutalebi and Green 2007).

2.2.2 Processing false friends

Interference effects for Full ILHs are quite well attested. For example, Dijkstra et al. (2000) used a go/no-go task and found inhibitory effects for ILHs even in the L1 version of the task. Martín et al. (2010) asked Spanish-English bilingual participants to decide whether pairs of L2 (English) words were semantically related. Word pairs consisted of an ILH paired with either an L2 word related to the L1 meaning, or to an unrelated L2 word. Participants responded significantly more slowly to the former condition, indicating that the ILHs were activating both word meanings, resulting in interference.

Such findings for Full ILHs are in full contrast to the facilitative effects robustly attested for cognates. Evidence for a "cognate advantage" has come from a several studies using a variety of experimental tasks. For example, compared to non-cognate translation equivalents, it has been found that cognates are recognized faster (e.g. Comesaña et al. 2012); responded to more quickly in picture naming tasks (e.g. Costa et al. 2005); and translated faster in both directions (e.g. de Groot et al. 1994).

If cognates have a processing advantage for bilinguals and ILHs have a processing disadvantage, this suggests that for Partial ILHs, these opposing forces are both present. Indeed, some studies have found a multiple translation disadvantage for cognates. For example, using a lexical decision task in the L2, Sunderman and Schwartz (2008) found that reaction times were faster for cognates than for Partial ILHs, but that both of these word types were responded to more quickly to than were non-cognates. Further, a study by Boada et al. (2012) used a translation recognition task in both directions (L1-to-L2 and L2-to-L1) and found that both non-cognates and cognates were negatively impacted

by having multiple translations. These results suggest that Partial ILHs benefit from their cognate status, but that this advantage is attenuated by having multiple translations.

Studies examining cross-language ambiguity with *non-cognate* translations have also found a multiple translation disadvantage for L2 learners. For example, Degani and Tokowicz (2010) found that both "meaning-ambiguous" translations (e.g. English *size*; Dutch *grootte/maat* "size") and "form-ambiguous" translations (e.g. English *change*; Dutch *verandering* "alteration" and *wisselgeld* "small coins") posed significant difficulties for L2 learners of Dutch. Interestingly, this distinction mirrors that between Partial FCs, which have a high level of semantic overlap, and Partial ILHs, which do not. It is therefore notable that the "meaning-ambiguous" translations were found to be particularly difficult for L2 learners. This suggests that the additional semantic overlap of Partial FCs may make them particularly difficult to acquire.

Finally, Full FCs represent a grey area between the complete semantic overlap of cognates and the complete lack of semantic overlap of Full ILHs. Further, because the degree of semantic overlap of FCs is variable from word pair to word pair, it is difficult to find a homogeneous group of stimuli to test. For example, both the English-French word pairs *demand-demande* (meaning "ask") and *rest-reste* (meaning "stay") have related but different meanings; however, does one word pair have more common properties than the other? This is perhaps the reason why these types of false friends have received virtually no attention in psycholinguistic investigations to date.

3. The current study

Given the issues regarding False Friends discussed above, more research is needed to understand the impact of cross-language lexical ambiguity on processing and lexical organization, and how this changes as a function of L2 proficiency. The study presented here aims to take the first step towards that objective. Specifically, it has three goals.

The first is to obtain similarity ratings from English monolingual speakers between French False Friend word meanings and those of their English False Friend counterparts. For example, for the false friend pair *EN: library, FR: librairie* ("bookstore"), the similarity rating for the English pair LIBRARY-BOOKSTORE was obtained. This was used to assess how closely related the meanings of false friends are perceived to be without the influence of L2 (French) knowledge. To do this, we tested a group of functional English monolinguals. It was expected that word meanings associated with FCs would be rated as being highly similar but not synonymous. On the other hand, it was expected that word meanings associated with ILHs would be rated as very dissimilar or completely different. These results were also used to confirm the classification of the stimuli items, and to serve as a baseline for comparison needed for the next goal.

The second goal is to determine whether, for bilinguals, L2 word meanings of false friends interfere in an L1-only task. To do this, a group of English-French bilinguals were tested using the same task, and their results were compared to those of the functional monolinguals. It was expected that the bilinguals' ratings of the critical word pairs would be significantly higher due to interference from the L2 False Friend.

Finally, it was hypothesized that the amount of interference would be dependent on L2 proficiency, such that the more proficient bilinguals would either experience more interference because of their stronger L2 mental representations, or less interference due to better language control skills. To investigate this, we divided the bilingual participants into two groups based on their L2 proficiency and compared their interference effects.

4. Methods

4.1 Participants

The data of 39 participants (35 female) were included in the study. All were between the ages of 18 and 22 (mean=19.3; SD=0.85) and lived in the Ottawa-Gatineau region at the time of participation. They were native speakers of Canadian English with current self-reported proficiency rated as either "native" or "near-native". All participants reported at least some exposure to French. None had language proficiency in another language exceeding that of English or French.

For 17 of these participants, exposure to French was solely in the context of a Core French elementary school program. As a group, these participants reported a mean self-rated French proficiency of 6.8/20[2] (SD=2.7) and a mean French cloze task score (see Section 4.4) of 5.5% (SD=5.2%). As such, these participants were considered to be the "Functional Monolingual" control group, even though they had this exposure to French.

The remaining 22 participants, the "bilinguals", were divided into two groups based on their L2 proficiency: High-Proficiency and Mid-Proficiency. The cut-off point used to separate these two groups was 50% on the French cloze task (see Section 4.4 for details).

Table 1 presents the characteristics of each participant group. An independent samples t-test shows that the two bilingual groups differed significantly in terms of self-rated L2 proficiency ($t(20)=2.76$, $p=.012$). Also, the participants' self-rated proficiency scores for French were significantly correlated with their cloze task score ($r=.857$, $p<.001$), demonstrating that these two L2 proficiency measures were in good agreement.

Table 1. Language backgrounds of participant groups (means with standard deviations).

	Functional Monolinguals (n = 17)	**Mid-Proficiency Bilinguals** (n = 12)	**High-Proficiency Bilinguals** (n =10)
Age at testing	19.3yrs (0.7yrs)	19.2yrs (0.9yrs)	19.3 yrs (1.1yrs)
Age of first exposure	6.5yrs (2.8yrs)	3.5yrs (2.7yrs)	2.4 yrs (3.5yrs)
Age of L2 immersion	N/A	3.8yrs (2.6yrs)	4.6 yrs (6.8yrs)
Self-rated proficiency L2 (Fr):	6.8/20 (2.7)	13.4/20 (3.2)	17.1/20 (3.0)
Self-rated proficiency L1 (En):	19.7/20 (1.0)	19.8/20 (0.6)	19.5/20 (1.1)
Cloze task score (L2)	5.5% (5.2%)	41.5% (4.1%)	66.2% (11.4%)

4.2 Stimuli

Critical stimuli were English word pairs consisting of an English false friend (Word A) and the English translation of the corresponding French false friend (Word B). Only open-class words were used. See Table 2 for examples of each stimuli type.

In terms of form, Word A and its corresponding French false friend were to have a significant amount overlap, whereas Word A and Word B were to have very little. Because the experimental task involved visual word recognition, only orthographic overlap was objectively measured. This was done using the MatchCalc program (see

[2] The total score is based on a scale from 1 (very low) to 5 (native) in four skill areas (reading, writing, speaking, listening), resulting in a maximum possible score of 20 and a minimum possible score of 4.

Table 2. Examples of stimuli items.

	False Friend Pair (English, French)	**WORD A:** (English False Friend)	**WORD B:** (English Translation of French False Friend)
Full FC (n = 15)	library, librairie	library	bookstore
Full ILH (n = 12)	home, homme	home	man
Partial FC (n = 13)	corpse, corps	corpse	body
Partial ILH (n = 14)	peach, pêche	peach	fishing

Davis and Bowers (2006) for details). It was also required that there be at least one orthographic feature differentiating Word A and its corresponding French false friend, for example an additional accent (e.g. EN: reunion, FR: réunion), or transposed letters (e.g. EN: ponder, FR: pondre). This was important as it ensured that the experimental task explicitly involved only English words, as opposed to words that could be read as either English or French. The average orthographic overlap for each stimuli type, as well as their average lexical frequency[3], are shown in Table 3.

Table 3. Properties of each stimuli type (after exclusions/reclassifications of Section 5.1).

Stimuli Type	**Word A** (English False Friend)	**Word B** (Translation of French False Friend)		**French False Friend**	
	Freq. (wpm)	Freq. (wpm)	Orth. Overlap with Word A	Freq. (wpm)	Orth. Overlap with Word A
Full FCs	34.9	233.2	21.2%	78.9	75.9%
Full ILHs	101.4	300.8	13.4%	109.6	77.8%
Partial FCs	25.8	118.9	16.7%	94.9	82.9%
Partial ILHs	183.3	67.0	14.8%	146.2	70.4%

The *Canadian Oxford Dictionary* and the dictionary from the software program *Antidote* (http://www.antidote.info) were consulted in order to ensure that false friends were adequately classified as Full or Partial, depending on the number of meanings the French false friend and its English form-related counterpart (Word A) shared. Full FCs and ILHs were to each have a single dominant meaning in each language, whereas Partial FCs and ILHs were to have two dominant meanings in French, such that one translation was the form-related Word A, and the other was the form-unrelated Word B.

Control stimuli consisted of 101 non-cognate English word pairs with minimal orthographic overlap. Of these, 29 were used as practice stimuli and so were not included in the analysis. Crucially, they varied in terms of their degree of semantic overlap, from completely unrelated to synonymous. To classify these word pairs along the six-point scale, they were rated independently by the author and two research assistants with native English proficiency. Word pairs were only selected if their ratings fell within two points of the scale for all three researchers, and average ratings were used when there were such discrepancies. Pairs rated as a 6 were classified as "Synonymous"; those rated as 4-5 as "High Similarity"; those rated as 2-3 as "Low Similarity"; and those rated as 1 as "No Similarity". There were approximately the same number of word pairs in each category.

[3] Frequency counts for English words are from the $SUBTLEX_{US}$ database (Brysbaert and New 2009), and those for the French false friends are from the Lexique database (New et al. 2001).

Critical and control stimuli were combined into a single list of 155 items, and this list was divided into seven blocks of 21 items each, plus a practice block of eight items. The first block contained 21 control items, and each of the subsequent six experimental blocks contained 12 control and nine critical stimuli, both distributed as equally as possibly with respect to stimuli type. The order of presentation within each block was semi-randomized such that no two items of the same type appeared consecutively, and such that the first item in each block was always a control item.

4.3 Procedure

The experiment was designed using the *Presentation* stimulus delivery and experimental control program (Neurobehavioral Systems). Each trial consisted of a blank screen presented for 600 ms, a fixation cross presented for 1000 ms, and a word pair presented for 3000 ms in lower case letters, for a total trial duration of 4600 ms. The two words of each pair were separated by six spaces. Fixation crosses and word pairs were centered on the screen and were presented in black, 30 point Arial font on a light grey background. Responses were recorded using a button box, with the six horizontally arranged buttons representing the scale from 1 = "no similarity" (left) to 6 = "synonymy" (right).

Participants were tested in a sound-attenuated room. Before beginning, they were presented with written instructions on the computer screen, as well as verbal instructions by the experimenter. They were instructed to use the button box to rate how much in common the two words of each pair had, and were encouraged to answer as quickly and as accurately as possible. It was explicitly stated that all words would be in English. Between each block, participants were instructed to take a self-regulated break.

4.4 Other testing materials

Participants were required to complete three additional language assessment measures. The first, a language background questionnaire, was administered in order to assess eligibility and to characterize participants in terms of the information presented in Table 1. Secondly, participants completed a French cloze test (Tremblay 2011) in order to objectively assess their L2 proficiency and to divide the bilinguals into High- vs. Mid-proficiency groups. This test is considered to be a valid, reliable and practical method of obtaining a global assessment of L2 knowledge.

Finally, a written Translation Post-Test was administered in order to determine whether the false friends in question were adequately known to the bilinguals. It was also used to determine whether Mid- and High-proficiency bilinguals had similar knowledge of the False Friends in question. If so, then any observed differences in the interference effects between the two groups in the semantic relatedness task could be attributed to differences in L2 activation and not simply differences in underlying knowledge. This task involved translating the French targets into English and then evaluating their confidence in each response, from 1 (a guess) to 4 (certain). All 54 French false friends were tested, with the 28 Full FCs and ILHs in one list, and the 28 Partial FCs and ILHs in another. An additional 28 high frequency French words were added to the first list as filler. Participants were asked to produce both translations for Partial FCs and ILHs, and were told that one would have similar orthography to the French word, and the other not.

Participants completed the language background questionnaire before they performed the semantic relatedness task. Importantly, they always completed the French

cloze test and the Translation Post-Test afterwards. This order was chosen to minimize participants' likelihood of noticing that knowledge of French was relevant to the task; ideally, participants would be performing the experiment in "monolingual mode".

5. Results

5.1 Exclusions and reclassifications

The responses to the Translation Post-test for the ten most proficient participants were examined in order to determine whether any items should be excluded. This was done if all three of the following criteria were met: 1) three or fewer participants produced it; 2) the responses were highly variable (three or more different responses); and 3) certainty was rated low (2 or less) by five or more participants. Ten items were excluded based on these criteria: three Full FCs, two Full ILHs, two Partial FCs, and three Partial ILHs.

In order to determine whether the stimuli items were appropriately classified in terms of their amount of semantic overlap, the semantic relatedness task results for the Monolingual group were submitted to boxplots. When an outlier was identified, that item was reclassified into a more appropriate category, and the boxplots were rerun. If the item remained an outlier, it was excluded. If not, it was reclassified into this new group. This procedure was repeated until there were no outliers. Using this method, two Partial FCs were reclassified as Partial ILHs, one Partial ILH was excluded, and eight control items were reclassified (six from Synonymous to High Similarity, one from Low to High similarity, and one from High to Low similarity). The seemingly drastic change in the 'Synonymous" category was not entirely unexpected, as true synonymy does not really exist within a language. The participants tested here seemed sensitive to this fact.

The data from one of the 39 participants were excluded from the final analyses based on their performance in the semantic relatedness task. This participant used four of the six buttons less than 10% of the time each, leaving buttons two buttons used a combined 83% of the time (37% and 46%, respectively). This response distribution suggested that the participant was treating the semantic judgement as essentially binary.

5.2 Data

The data reported and analyzed in the following sections consist of the semantic ratings for the 38 participants and 43 items remaining after the above exclusions. For all analyses reported, statistical significance was evaluated at the 95% confidence interval and all planned pairwise and post-hoc comparisons used the Bonferroni correction.

It is important to note, however, that the results of the Translation Post-test indicated that the High-proficiency bilinguals *did* know more of the False Friend translations than did the Mid-proficiency bilinguals[4]. Further, it was also demonstrated that many of the functional monolinguals also knew some of the translations, though of course significantly fewer than did the two bilingual groups. To take these factors into account, analyses of the semantic relatedness task were conducted over two different data sets. Data Set 1 included all of the 44 critical items for each participant. Data Set 2 excluded translations that were *not* accurately produced in the post-test for each bilingual,

[4] A repeated-measures ANOVA indicated that Full and Partial ILHs in particular were better known to the high-proficiency group. Regression analyses revealed a significant correlation for all but the Full FCs.

and excluded translations that *were* accurately produced in the post-test for each monolingual. As this second data set may actually underestimate bilingual participants' actual knowledge, and also has lower statistical power due to fewer stimuli items, it was only used to determine whether or not the patterns found using Data Set 1 held true when the data was "controlled" for participants' apparent knowledge of the false friends in question. Generally, the two data sets were in agreement. As such, the results using Data Set 1 will be reported below, with those of Data Set 2 only included where relevant.

5.3 Monolingual baseline evaluations

Using the data for the 16 functional monolinguals, semantic similarity ratings for each type of stimuli were averaged across participants and across items (see Table 4). Participant and items analyses were conducted using one-way repeated measures ANOVA, with *Stimuli Type* as the repeated measure (8 levels: Synonymous, High Similarity, Low Similarity, No Similarity, Full FCs, Partial FCs, Full ILHs, Partial ILHs).

Table 4. Mean semantic similarity ratings (out of 6) for Monolinguals (n = 16).

	Participant Mean (SD)	**Item Mean (SD)**
Full FC (e.g. librairie, bookstore)	3.98 (0.60)	3.97 (0.88)
Partial FC (e.g. corpse, body)	4.59 (0.67)	4.58 (0.37)
Full ILH (e.g. avocado, lawyer)	2.03 (0.63)	2.04 (0.72)
Partial ILH (e.g. peach, fishing)	1.62 (0.59)	1.62 (0.25)
Control: Synonymous	5.38 (0.36)	5.38 (0.16)
Control: High similarity	4.50 (0.55)	4.51 (0.57)
Control: Low similarity	2.37 (0.65)	2.40 (0.69)
Control: No similarity	1.45 (0.57)	1.45 (0.24)

The main effect of *Stimuli Type* was significant ($F_1(7,105)=164.44$, MSE=37.90, $p_1<.001$; $F_2(7,107)=110.32$, MSE=33.23 $p_1<.001$). Planned pairwise comparisons were conducted to compare ratings between conditions. All four control conditions differed significantly from one another, with the Synonymous word pairs having the highest overall rating, followed by the High Similarity pairs, the Low Similarity pairs, and finally, the No Similarity pairs (all *p*'s<.002). This was as expected, particularly given that some items were reclassified.

In terms of the critical conditions, Full FCs were rated significantly higher than both Full ILHs ($p_1<.001$, $p_2<.001$) and Partial ILHs ($p_1<.001$, $p_2<.001$), and did not differ significantly from Partial FCs in the items analysis ($p_1=.026$, $p_2=.375$). They were also rated significantly higher than the No Similarity ($p_1<.001$, $p_2<.001$) and the Low Similarity ($p_1<.001$, $p_2<.001$) control word pairs, significantly lower than the Synonymous word pairs ($p_1<.001$, $p_2<.001$), but not significantly different from the High Similarity word pairs in the items analysis ($p_1=.016$, $p_2=.182$).

Partial FCs displayed the same pattern. They were rated significantly higher than the Full ILHs ($p_1<.001$, $p_2<.001$), the Partial ILHs ($p_1<.001$, $p_2<.001$, the No Similarity word pairs ($p_1<.001$, $p_2<.001$), and the Low Similarity word pairs ($p_1<.001$, $p_2<.001$), and were rated significantly lower than the Synonymous word pairs ($p_1=.001$, $p_2=.037$), but not significantly different from the High Similarity word pairs ($p_1=1$, $p_2=1$).

Full and Partial ILHs displayed the opposite pattern from the FCs. Full ILHs were rated significantly higher than both Full FCs (p_1<.001, p_2<.001) and Partial FCs (p_1<.001, p_2<.001), and did not differ significantly from Partial ILHs in the items analysis (p_1=.014, p_2=1). They were rated significantly lower than the Synonymous (p_1<.001, p_2<.001) and the High Similarity (p_1<.001, p_2<.001) control word pairs, significantly higher than the No Similarity word pairs in the participants analysis (p_1=.009, p_2=.207), but not significantly different from the Low Similarity pairs (p_1>.483, p_2=1).

Partial ILHs were rated significantly higher than both Full FCs (p_1<.001, p_2<.001) and Partial FCs (p_1<.001, p_2<.001. They were also rated significantly lower than the Synonymous (p_1<.001, p_2<.001) and the High Similarity (p_1<.001, p_2<.001) control word pairs, significantly higher than the Low Similarity word pairs (p_1=.005, p_2>.007), but not significantly different from the No Similarity word pairs (p_1=1, p_2=1).

To summarize, the patterns identified here for the critical conditions are largely as expected. The meanings of Full and Partial FCs were not treated as being synonymous. Partial FCs were rated the same as the Highly Similar control word pairs, while the Full FCs were rated slightly lower than the High Similarity control word pairs, but still in the upper half of the similarity scale and significantly higher than the Low Similarity control word pairs. The ratings of Full and Partial ILHs word pairs, on the other hand, indicate that they really do have little meaning in common.

5.4 Bilinguals

In order to determine whether, for bilinguals, L2 word meanings of false friends interfere in an L1-only task, the semantic ratings of the two bilingual groups were compared to those of the Functional English Monolinguals. If bilinguals rated the critical items higher on the similarity scale than did Monolinguals, this would be taken as evidence of L2 interference. The interference effects of High- and Mid-proficiency bilinguals were compared using both ANOVA and linear regression, the latter reflecting the fact that L2 proficiency is a continuous variable.

5.4.1 Group analyses

Using the data for all participants, divided into their proficiency groups, the semantic similarity ratings for each type of stimuli were averaged across participants and across items. The participants analysis was conducted using a two-way repeated measures ANOVA, with *Proficiency Group* as the between-participant variable (High, Mid and Monolingual), and *Stimuli Type* as the repeated measure (8 levels). The items analysis was conducted using a two-way repeated measures ANOVA, with *Proficiency Group* as the repeated measure and *Stimuli Type* as the between-items variable.

The main effect of *Proficiency Group* was not determined to be significant by participants but it was by items (F_2(2,214)=22.55, MSE=3.53, p_2<.001. This was likely due to the higher ratings of critical items by bilinguals. The main effect of *Stimuli Type* was determined to be significant by both participants (F_1(7,245)=246.27, MSE=65.85, p_1<.001) and items (F_2(7,107)=94.90, MSE=85.89, p_2<.001). As expected, the patterns across Proficiency groups mirrored those found for the Monolingual group.

Most importantly, the interaction between *Proficiency* and *Stimuli Type* was determined to be significant by both participants (F_1(14,245)=5.17, MSE=1.38, p_1<.001) and items (F_2(14,214)=6.69, MSE=1.04, p_2<.001). These results are shown in Figure 1. In

order to investigate this interaction, eight separate one-way ANOVA were conducted for each *Stimuli Type*, with *Proficiency Group* as the independent variable. For the participants analysis, *Proficiency Group* was a between-participants variable, whereas for the items analysis, it was a within-item repeated measure. For Full FCs, a significant main effect was found ($F_1(2,35)=5.90$, MSE=1.92, $p_1=.006$; $F_2(2,22)=9.00$, MSE=1.94, $p_2=.001$). Planned pairwise comparisons indicated that the High-proficiency bilinguals rated them as significantly higher than did the Monolinguals (difference=0.74, $p_1=.008$, $p_2=.013$), and the Mid-proficiency bilinguals rated them marginally higher than did the Monolinguals (difference=0.52, $p_1=.071$, $p_2=.078$). For Full ILHs, a significant main effect was also found ($F_1(2,35)=9.27$, MSE=3.61, $p_1=.001$; $F_2(2,18)=7.95$, MSE=2.76, $p_2=.003$). Again, planned pairwise comparisons indicated that the High-proficiency bilinguals rated them as significantly higher than did Monolinguals (difference=0.85, $p_1=.005$, $p_2=.016$) and the Mid-proficiency bilinguals rated them marginally higher than did the Monolinguals (difference=0.91, $p_1=.002$, $p_2=.088$). For Partial FCs, the main effect was not determined to be significant ($F_1(2,35)=.175$, MSE =.084, $p_1=.840$; $F_2(2,16)=1.14$, MSE=.131, $p_2=.345$. This was not expected. For Partial ILHs, the main effect was again found to be significant ($F_1(2,35)=9.41$, MSE=4.82, $p_1<.001$; $F_2(2,22)=9.96$, MSE=3.99, $p_2=.001$. Planned pairwise comparisons indicated that the High-proficiency bilinguals rated them significantly higher than did the Monolinguals (difference=1.17, $p_1=.001$, $p_2=.003$) and the Mid-proficiency bilinguals rated them marginally higher than did the Monolinguals (difference=0.84, $p_1=.012$, $p_2=.081$).

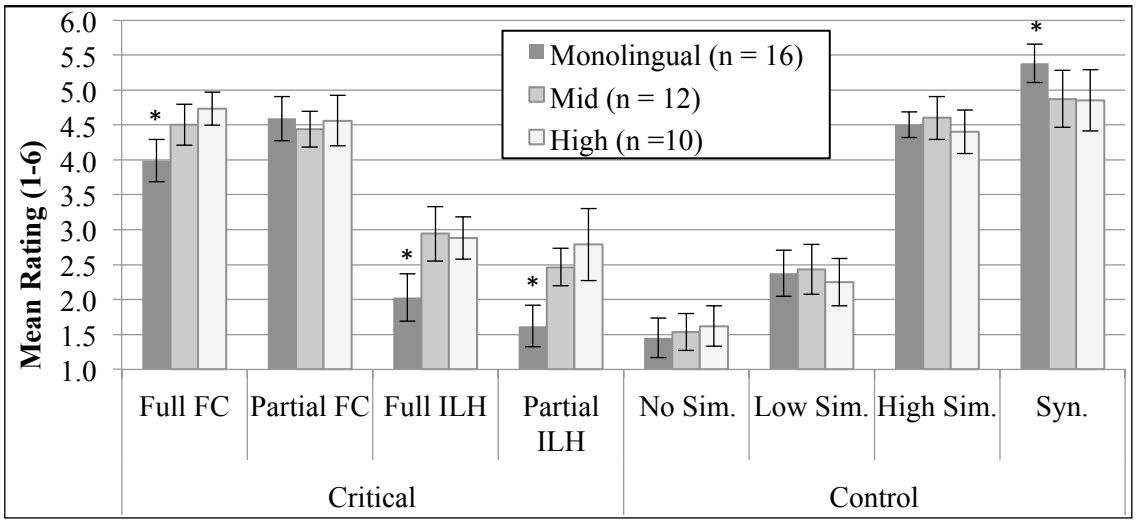

Figure 1. Mean ratings for each stimuli type by Proficiency Group using Data Set 1. Error bars represent the standard deviation.

To summarize, results for three of the four critical conditions exhibited the same pattern: the bilingual groups rated them significantly higher than did the monolinguals, but there was no significant difference between the two bilingual groups. The exception to this pattern was the Partial FCs, where all three groups rated them similarly. While this was unexpected, it can be explained as a ceiling effect; the Partial FCs were rated quite high on the similarity scale by Monolinguals (see Table 4 in Section 5.3), and so perhaps they could not be rated any higher without creeping into "synonymous" territory.

The control conditions generally exhibited the expected patterns. For the No Similarity pairs, the main effect of Proficiency Group was not found to be significant by participant, though it was determined to be marginally significant by item ($F_2(2,34)=3.14$, MSE=.121, p_2=.056). Post-hoc tests suggested that these items were rated significantly higher by the High-proficiency group compared to the Monolinguals (difference=0.16, p_2=.047). Given the small difference in magnitude and the marginal *p*-values, this could simply be a Type I error. For the Low Similarity pairs, the main effect was neither significant by participant nor by item means. For the High Similarity pairs, this effect was not significant in the participants analysis, but was marginally significant in the items analysis ($F_2(2,46)=2.84$, MSE=.29, p_2=.069); however, post-hoc tests indicated that the groups did not differ significantly from one another (all p_2's>.15). Finally, for the Synonymous pairs, the main effect was determined to be significant in the participants analysis ($F_1(2,35)=4.48$, MSE=1.24, p_1=.018), but not in the items analysis. Post-hoc tests showed that the High- and Mid-proficiency bilinguals rated the similarity of these pairs significantly lower than did the Monolinguals (difference=0.53, p_1=.052; and difference=0.51, p_1=.049, respectively). These results indicate that the Monolinguals and Bilinguals were rating the control items in a similar way, with the exception of the Synonymous pairs, for which both groups of bilinguals rated them significantly lower than did the Monolinguals.

5.4.2 Regression analyses

A linear regression was conducted over the mean semantic ratings of each of the eight *Stimuli Types*, with *Cloze Task Score* as the predictor variable. The results from Data Set 1 for all 38 participants largely mirror those of the Group Analyses: the regression models are significant for the Full FCs, the Full ILHs and the Partial ILHs (all r^2>.242, all *p*'s <.003) but not for the Partial FCs (r^2=.003, *p*=.765). Using Data Set 2, however, the results are actually stronger (see Table 5). Here, the regression model for all Stimuli Types reaches significance. This constitutes stronger evidence that the interference experienced by the bilinguals for the Partial FCs was obscured by a ceiling effect.

Table 5. Correlations between ratings and Cloze Task score using Data Set 2 (N = 38).

Stimuli Type	**Correlation (r)**	**r^2**	**Model**
Full FCs	.489	.239	F(1,36) = 11.309, MSE = 5.606, *p* = .002
Full ILHs	.537	.289	F(1,36) = 14.609, MSE = 7.277, *p* = .001
Partial FCs	.401	.161	F(1,36) = 6.911, MSE = 5.434, *p* = .013
Partial ILHs	.703	.494	F(1,36) = 35.115, MSE = 20.602, *p* < .001
No Similarity	.132	.017	F(1,36) = .634, MSE = .194, *p* = .431
Low Similarity	.090	.008	F(1,36) = .292, MSE = .133, *p* = .592
High Similarity	.023	.001	F(1.36) = .019, MSE = .010, *p* = .891
Synonymous	.416	.173	F(1,36) = 7.542, MSE = 2.121, *p* = .009

It is also important to note that when the data of Monolinguals are excluded, none of the correlations or regression models reached significance, using either Data Set 1 or 2. While this could be reflective of the small sample size, it suggests that both groups of bilinguals experienced similar levels of L2 interference from these particular stimuli.

6. Discussion

The main findings of the study presented here are that covert L2 meanings of L1 false friends cause interference even when the task was conducted in the L1, and that this was true for all types of false friends, and for bilinguals with different L2 proficiency levels.

These findings are consistent with those of previous studies investigating Full ILHs, e.g. those of Martín et al. (2010), and particularly those of Dijkstra et al. (2000), who also found interference effects from the L2 to the L1. The results of the current study further extend these effects along the "semantic overlap continuum" to Full FCs. These results also demonstrate that the within-language semantic ambiguity of both Partial ILHs and Partial FCs is processed. These results are in line with those from previous studies investigating Partial ILHs, for example those of Boada et al. (2012) and Sunderman and Schwartz (2008), who found that cognates were negatively impacted by having multiple translations. The current study shows that this is so even when the task is performed in the L1, and extends the results to include Partial FCs.

The current study also suggests that complete orthographic overlap is not necessary to find cross-language effects for false friends, as is the case for cognates (e.g. Lemhöfer and Dijkstra 2004, Voga and Grainger 2007, Duñabeitia et al. 2010). Cross-language word pairs with incomplete orthographic overlap have been used in many studies investigating cognates (e.g. de Groot and Nas 1991, van Hell and Dijkstra 2002), but not in those investigating ILHs, where orthographic overlap (but not necessarily phonological overlap) is usually total. This supports connectionist models of the bilingual lexicon such as the BIA models (Dijktra and van Heuven 1998, 2002) that assume a shared lexical system at lower levels of representation (e.g. orthography).

Further, it was also hypothesized that one group of bilinguals would experience less L2 interference than the other; either the Mid-Proficiency group because they have weaker L2 mental representations, or the High-Proficiency group because they possess stronger cognitive control skills that enable them to better suppress activation of their other language. Neither of these scenarios, however, was found to be the case. It is of course possible that these differences do exist, but that the methods used here were not able to detect them. First of all, the sample size was quite small for correlation and regression analyses, and so a lack of statistical significance could simply be a reflection of this. Secondly, it is possible that the L2 proficiency measure used may not have been perfectly suited to the specific goals of the study. Cloze tests are used to obtain a global assessment of L2 knowledge by quickly and simultaneously testing morphosyntactic, lexical, and discourse competence (Tremblay 2011); however, this may be too general for the current purposes, as it tells us more about *what* participants know explicitly as opposed to *how well* they know it implicitly. Using a proficiency measure that focuses on quality and depth of lexical knowledge might be a better test. Alternatively, examining the impact of other variables such as frequency of use and manner of acquisition (i.e. naturalistic/instructional) might be better proxies for quality and depth of L2 knowledge.

Finally, the semantic relatedness task used here may have lacked the covertness needed to truly tap into implicit lexical processing. It is important to note that all bilingual participants noticed the connection between the stimuli items and the underlying false friends being tested. Both languages were therefore consciously activated even though participants were not led to believe, explicitly or implicitly, that any L2 knowledge was relevant. An indirect consequence of this might have been bilinguals' lower ratings for the Synonymous control items; for example, when a

bilingual participant sees the English word pair AVOCADO-LAWYER and intuitively rates it high on the semantic scale as a result of interference from the French word "avocat", they might immediately realize that they had made a mistake, leading them to be suspicious of subsequent word pairs that seem to be highly similar in meaning. That being said, the bilinguals were not able to suppress this L2 knowledge quickly enough to strategically limit interference in this speeded task, as such effects were indeed found.

In order to truly test subconscious L2 effects, a more implicit task is needed. A good candidate for such a task is a lexical decision task with masked priming (Forster and Davis 1984). Importantly, a future study employing this methodology could make use of the stimuli items developed and tested in the study just presented. Indeed, the results for the Monolingual group indicate that the stimuli selected for each of the false friend types are now properly classified and form distinct categories. It must be said, however, that there are relatively few items in each category: between 9 and 12. This small number reflects the challenge of finding false friends that are suitable for testing. Further, those items that were selected in the end are still less than ideal. For example, it was not possible to control for their relative frequency of occurrence in each of their respective languages without further reducing the number of usable stimuli items. It does not seem, however, that frequency effects impacted the results of this study. Table 3 in Section 4.2 shows that, for Full FCs, Full ILHs and Partial FCs, Word A (the English false friend, e.g. *library*) had a lower average frequency than Word B (the English translation of the French false friend, e.g. *bookstore*). The opposite was true for Partial ILHs. Because interference effects were still found for all types of false friend despite this asymmetry, this suggests that relative frequency was not an important factor.

7. Conclusions

False friends are a complex subset of vocabulary items for bilinguals to both learn and process, and the current study aimed to make preliminary but comprehensive steps to better understand them. While the results have contributed additional evidence for false friend interference effects and extend previous findings in new directions, they can shed little light on the question of how these different types of false friends are represented and organized mentally, due to the not-so-subconscious nature of experiment used. Understanding how such words are cognitively organized, represented and processed is not only of interest to those investigating the bilingual mental lexicon from a theoretical standpoint, but also to those teaching in the L2 classroom, as they directly address the challenges that their students face when acquiring such deceiving vocabulary items. As such, it is an important area of investigation, and the stage has been set for future work.

References

Abutalebi, Jubin and David Green. 2007. Bilingual language production: the neurocognition of language representation and control. *Journal of Neurolinguistics*, 20:242–275.

Boada, Roger, Rosa Sánchez-Casas, José Maria Gavilán, José Eugenio García-Albea, and Natasha Tokowicz. 2012. Effect of multiple translations and cognate status on translation recognition performance of balanced bilinguals. *Bilingualism: Language and Cognition* 16:183–197.

Brysbaert, Marc, and Boris New. 2009. Moving beyond Kucera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods* 41:977–990.

Comesaña, Monserrat, Ana Paula Soares, Rosa Sánchez-Casas, and Cátia Lima. 2012. Lexical and semantic representations in the acquisition of L2 cognate and non-cognate words: Evidence from two learning methods in children. *British Journal of Psychology* 103:378–392.
Costa, Albert, Mikel Santesteban, and Agnès Caño. 2005. On the facilitatory effects of cognate words in bilingual speech production. *Brain and Language* 94:94–103.
Davis, Colin J., and Jeffrey S. Bowers. 2006. Contrasting five different theories of letter position coding: Evidence from orthographic similarity effects. *Journal of Experimental Psychology: Human Perception and Performance* 32:535–557.
Degani, Tamar, and Natasha Tokowicz. 2010. Ambiguous words are harder to learn. *Bilingualism: Language and Cognition* 13:299–314.
Dijkstra Ton, and Walter J. B. van Heuven. 1998. The BIA model and bilingual word recognition. In *Localist connectionist approaches to human cognition*, ed. Jonathan Grainger and Arthur M. Jacobs, 189–225. Mahwah, NJ: Lawrence Erlbaum.
Dijkstra, Ton, and Walter J. B. van Heuven. 2002. The architecture of the bilingual word recognition system: From identification to system. *Bilingualism: Language and Cognition* 5:175–197.
Dijkstra, Ton, Mark P.H. Timmermans, and Herbert J. Schriefers. 2000. On being blinded by your other language: Effects of task demands on interlingual homograph recognition. *Journal of Memory and Language* 42:445–464.
Dimitropoulou, Maria, Jon Andoni Duñabeitia, and Manuel Carreiras. 2011. Masked translation priming effects with low proficient bilinguals. *Memory and Cognition* 39:260–275.
Duñabeitia, Jon Andoni, Manuel Perea, and Manuel Carreiras. 2010. Masked translation priming effects with highly proficient simultaneous bilinguals. *Experimental Psychology* 57:98–107.
Finkbeiner, Matthew, Kenneth I. Forster, Janet Nicol, and Kumiko Nakumura. 2004. The role of polysemy in masked semantic and translation priming. *Journal of Memory and Language* 51:1–22.
Forster, Kenneth I., and Chris Davis. 1984. Repetition priming and frequency attenuation in lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 10:680–698.
de Groot, Annette M. B., and G. L. Nas. 1991. Lexical representation of cognates and noncognates in compound bilinguals. *Journal of Memory and Language* 30:90–123.
de Groot, Annette M. B., Lucia Dannenburg, and Janet G. van Hell. 1994. Forward and backward word translation by bilinguals. *Journal of Memory and Language* 33:600–629.
van Hell, Janet, and Ton Dijkstra. 2002. Foreign language knowledge can influence native language performance. *Psychonomic Bulletin and Review*, 9:780–789.
Jiang, Nan. 1999. Testing processing explanations for the asymmetry in masked cross-language priming. *Bilingualism: Language and Cognition* 2:59–75.
Kroll, Judith F., and Erika Stewart. 1994. Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language* 33:149–174.
Lemhöfer, Kristin, and Ton Dijkstra. 2004. Recognizing cognates and interlingual homographs: Effects of code similarity in language-specific and generalized lexical decision. *Memory and Cognition* 32:533–550.
Martín, María Cruz, Pedro Macizo, and Maria Teresa Bajo. 2010. Time course of inhibitory processes in bilingual language processing. *British Journal of Psychology* 101:679–693.
New, Boris, Christophe Pallier, Ludovic Ferrand, and Raphael Matos. 2001. Une base de données lexicales du français contemporain sur internet: LEXIQUE, *L'Année Psychologique* 101:447–462.
Potter, Mary C., Kwok-Fai So, Barbara Von Eckardt, and Laurie B. Feldman. 1984. Lexical and conceptual representation in beginning and more proficient bilinguals. *Journal of Verbal Learning and Verbal Behavior* 23:23–38.
Sabourin, Laura, Christie Brien, and Michèle Burkholder. 2014. The effect of age of L2 acquisition on the organization of the bilingual lexicon: Evidence from masked priming. *Bilingualism: Language and Cognition* 17(3):542–555.
Sunderman, Gretchen, and Ana I. Schwartz. 2008. Using cognates to investigate cross-language competition in second language processing. *TESOL Quarterly* 42:527–536.
Swinney, David. 1979. Lexical access during sentence comprehension: (Re)consideration of context effects. *Journal of Verbal Learning and Verbal Behaviour* 18:645–659.
Tremblay, Annie. 2011. Proficiency assessment standards in second language acquisition research: "Clozing" the gap. *Studies in Second Language Acquisition* 33:339–372.