Cross-language similarity in L2 processing: An eye-tracking study

by

Alba Tuninetti

B.A., University of Pennsylvania, 2009

Submitted to the Graduate Faculty of

Kenneth P. Dietrich School of Arts and Sciences in partial fulfillment

of the requirements for the degree of

Master of Science

University of Pittsburgh

2012

UNIVERSITY OF PITTSBURGH

Kenneth P. Dietrich School of Arts and Sciences

This thesis was presented

by

Alba Tuninetti

It was defended on

January 24, 2012

and approved by

Tessa Warren, Associate Professor, Department of Psychology

Charles A. Perfetti, Distinguished Professor, Department of Psychology

Thesis Director: Natasha Tokowicz, Associate Professor, Department of Psychology

Cross-language similarity in L2 processing: An eye-tracking study

Alba Tuninetti, M.S.

University of Pittsburgh, 2012

How do second language (L2) learners process L2 syntax? Using online (eye-tracking) and offline (grammaticality judgment) measures, this study examined how L2 learners of English process grammaticality in syntactic structures that were formed similarly, differently, or uniquely in L1 and L2. Native Arabic and native Mandarin Chinese speakers who were studying English as an L2 read grammatical and ungrammatical English sentences. Their reading behavior was compared to that of monolingual English speakers. L2 learners showed native-like processing of the syntactic violations compared to the native English speakers at some measures, but also demonstrated modulations in how sensitive they were to the violations based on similarity condition and cue strength. These results provide important considerations for models of L2 acquisition and bilingualism, such as the Unified Competition Model (MacWhinney, 2005) and the Shallow Structure Hypothesis (Clahsen & Felser, 2006).

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1.0 INTRODUCTION

Syntax varies widely across languages, from larger aspects such as phrase structure to more fine-grained differences such as morphosyntax. These variations are thought to influence second language (L2) processing and learning in different ways (e.g., Clahsen & Felser, 2006; MacWhinney, 2005). The present study builds on previous research on cross-language similarity in (morpho)syntax in L2 learners and expands it to purely syntactic constructions using a syntactic violation paradigm in an eye-tracking study. Native Arabic, native Mandarin Chinese, and native English speakers were tested on three types of sentences, two of which included syntactic violations that differed in their similarity across languages to test if L2 learners differentially process different types of syntactic violations and how they accomplish this.

Multiple theories have been put forward about how two syntactic systems may coexist in the mind of a bilingual. The two models most relevant to the present research are the Shallow Structure Hypothesis (Clahsen & Felser, 2006) and the Unified Competition Model (MacWhinney, 2005).

The Shallow Structure Hypothesis (SSH) put forth by Clahsen and Felser (2006a) addresses transfer and processing in two ways. First, it distinguishes between L2 morphosyntactic and syntactic processing. Specifically, it states that L2 learners have only shallow parsing of the L2 available to them because their L2 grammar does not provide the syntactic information needed to process complex structures in a native-like way. Thus, more complex structures will be resolved through "semantic, associative and surface information" rather than using syntactic cues (Clahsen & Felser, 2006a, p. 31). Second, the SSH posits that there is no positive syntactic transfer from L1 to L2; transfer occurs only at the word or morphosyntactic level (phonological, morphological, and semantic properties). As evidence for this model, the authors cite studies that have used offline measures to show L1 transfer and contend that studies that use online measures have had more difficulty in showing a clear result for transfer. The SSH sharply outlines the difference between shallow and deep parsing of an L2 and claims that L2 learners have only shallow representations of the L2 and cannot show native-like resolution at either early or late measures.

A study by Papadopoulou and Clahsen (2003) lends support to the idea that shallow processing predominates in complex L2 structures. In that study, highly proficient L2 Greek speakers (almost all had English as L1) did not show native-like ambiguity resolution preferences in syntactically ambiguous settings with online measures, even when the structures across languages were similar. Similarly, a study by Felser, Roberts, Gross, and Marinis (2003), showed that adult L2 speakers demonstrated lexically influenced disambiguation preferences in both offline (questionnaire) and online (self-paced reading) tasks for relative-clause attachment preferences in their L2, but did not apply the kinds of structurally-based resolution strategies that were seen in monolingual adults.

Interestingly, this kind of non-native like processing can be seen even in native speakers of the tested language. For example, Ferreira, Bailey, and Ferraro (2002) provided evidence that native speakers do not always have a syntactically rich representation of utterances. Using garden-path and passive sentences, they demonstrated that native speakers use what they term "good-enough" processing. This concept postulates that language comprehension sometimes is only partial and "good enough" to perform the task that the speaker is currently trying to finish (such as engaging in dialogue). Therefore, if native speakers do not fully parse sentences, it follows that L2 speakers may show the same pattern – they could show native-like processing for more lexically based structures, but show non-native like processing for more complex hierarchical structures, such as *wh*-questions and garden-path sentences. In this case, it is important to consider what the baseline is for the non-native speakers – if native speakers of a language do not always have a full, syntactically rich representation of a sentence, then non-native speakers should not be expected to exhibit this.

However, Clahsen and Felser (2006b) allow that native-like processing can occur for simpler syntactic constructions that rely on local and adjacent attachments. They state that adult L2 speakers rely on lexical, semantic, and pragmatic cues in the same way as native speakers, but this same reliability is absent when processing complex syntactic structures. Specifically, a learner's L2 grammar does not provide the necessary information to parse complex syntax (i.e., non-local dependencies such as *wh*-questions) in a native-like way. Assuming that having a sufficiently detailed and implicit knowledge of the L2 grammar is a prerequisite for successful processing, then non-native like processing may prevail with L2 learners because this knowledge is underspecified or absent. Therefore, even though the structure of the basic processing system is similar across languages, shallow processing will predominate for structures that lead to complex hierarchical structures. However, word-level processing and morphosyntactic feature matching can be more easily grasped as grammatical proficiency in the L2 increases, and can eventually become native-like (Clahsen & Felser, 2006b).

In contrast to the SSH, the Unified Competition Model (UCM; MacWhinney, 2005) states that "whatever can transfer will," and this transfer is a function of similarity between the two languages. Specifically, the UCM posits that structures that are *similar* across languages—realized in similar ways in L1 and L2—will share more common resources for processing; they may be easiest to learn during the acquisition of an L2 because of positive transfer from the L1 and a lack of competition between the two languages. *Different* structures – structures that are present in both languages but implemented differently at the surface level – lead to competition and therefore should be the hardest to learn and process. Finally, *unique* constructions—those that occur in L2 but not L1—should be easier to learn and process than the different constructions because there is no transfer but also no competition between the languages. Thus, learning of the structures is all that needs to happen. The relative ease of learning will depend on cues such as word order, subject-verb agreement, case marking, etc. These cues vary in strength based on the availability and reliability with which they appear in the input, and provide information about the function of a linguistic unit. For example, in the sentence "The dog bit the man", the cues of order of the nouns, *man* and *dog*, as well as prior world knowledge, provide information as to which is the agent and which is the patient. However, if the order is switched to "The man bit the dog", the cues are in conflict with each other because the order of the nouns is counter to what the reader expects and processing is slowed. This conflict is eventually resolved because the more-reliable cue prevails over the less-reliable cue (in this case, word order prevails over prior world knowledge). This demonstrates the importance of cue reliability during successful processing of linguistic phrases. To use cue strength as a strategy to learn differences and similarities between languages, contrastive forms need to be readily available, so that competition can highlight the differences between linguistic forms become apparent. For example, the cue of animacy is present and reliable only when there is an inanimate and an animate object in a sentence, contrasting the two forms, e.g. "The dog (*animate*) bit the eraser (*inanimate*)" (MacWhinney, 2005). MacWhinney and Pleh's (1988) study of Hungarian and Japanese speakers supports the UCM by showing that native speakers of these languages have trouble processing phrase structures (relative clauses) that differ between their L1 (Hungarian and Japanese, respectively) and their L2 (English for both groups), showing that contrastive cues present between L1 and L2 can cause processing difficulties during L2 comprehension.

The use of these contrastive cues also varies with L2 proficiency. McDonald (1987) examined how English-Dutch and Dutch-English bilinguals with varying levels of L2 exposure use cues to interpret native language sentences with dative constructions, relative clauses, and noun-verb-noun (NVN) sentences. The constructions differed in how similar or different they were between the two languages of the bilinguals across three experiments. The results demonstrated that depending on how similar or different the constructions were, there was transfer of cues from L1 to L2, and this transfer was modulated by the degree of participants' L2 exposure, such that increased L2 exposure correlated with using L2 cues for interpretation instead of transferring L1 cues (as was evident for the participants with less L2 exposure). McDonald proposed a model that used feedback from the environment (e.g., corrections, internal conflict) to adjust previously-learned cue strengths. In this model, the L2 learner assigns the weights of L1 cues to L2 input, so competition arises when the L1 cue and the correct L2 usage differ. With increasing L2 exposure and feedback, the cues are gradually readjusted to reflect the appropriate L2 rules. This model supports L2 learning as interplay between competition, cross-language similarity, and cue strength and availability.

Tolentino and Tokowicz (2011) addressed the SSH and UCM in detail in their review of (morpho)syntactic processing, investigating the effects of cross-language similarity in studies using ERP and functional magnetic resonance imaging (fMRI) methods. Structures that were cross-linguistically similar across languages showed comparable processing, but results varied more with cross-linguistically different or unique constructions. The main results across studies contradict some of the assumptions of the SSH because L2 speakers do seem to show early sensitivity to syntactic violations and to transfer syntax from their L1.

In a similar vein, Dussias (2003) showed that highly proficient Spanish-English speakers use relative clause attachment preferences that do not correspond to their native language in offline judgments. That is, native Spanish speakers who speak English proficiently may use English attachment preferences when interpreting relative clauses in Spanish. However, English-Spanish speakers showed clause attachment preferences that were consistent with L1. The two groups of participants did not have significant differences in proficiency level, but there were significant differences in how long they had lived in an L2-speaking country (i.e., native Spanish speakers in the USA and native English speakers in Spain). The author takes this into account when making her conclusions because the native language of the testing environment for all participants, English, may have influenced the participants such that English language cues were more available to the participants. Therefore, their judgments may have had more influence from the testing environment.

This evidence that even transfer from L2 to L1 is possible strongly supports models that allow transfer and cue noticing. Dussias (2003) speculated that the influence of testing language may have contributed to the results – because all of the speakers were in an English-speaking environment when they were tested, the English language cues may have been more readily available, so they showed judgments that conformed to the testing environment (Dussias, 2003). These findings are inconsistent with the predictions of the SSH because they show that L2 speakers may, in fact, demonstrate native-like parsing of syntactic structures if they are proficient enough and are able to make use of cues such as environmental constraints, to elicit syntactic resolution that belongs to their L2, and not their L1.

Previous research has also provided empirical evidence on how cross-language similarity affects L2 processing. In one study, Tokowicz and MacWhinney (2005) examined morphosyntactic structures (e.g., number agreement, gender agreement) as a function of cross-language similarity in an event-related potential (ERP) study. They focused on the P600 component, which is a positive-going wave seen in the presence of syntactic violations (more positive to ungrammatical than grammatical sentences) and syntactic complexity resolution during continuous electroencephalography (EEG) recordings from the scalp (e.g., Osterhout & Nicol 1999). Consistent with the UCM predictions, ERP responses showed more sensitivity (as measured by the P600 component) to similar and unique constructions than to the different construction. This, again, is most likely because the similar condition has the most transfer from L1 to L2 and the unique condition has no competition from L1.

Transfer effects have also been found in offline judgments of Dutch grammar. Sabourin, Stowe, and de Haan (2006) tested three language groups of L2 Dutch speakers; German-Dutch, English-Dutch, and Romance language-Dutch participants judged the gender of Dutch nouns. Although all participants scored very well in general with an average of 80% accuracy, there were significant differences between the language groups. Native English speakers performed the worst out of the three language groups, Romance-language groups followed, and the native German speakers had the best performance. Thus, the language group that has the most similar gender markings to Dutch (German) had the best performance. The group that has a gender system instantiated differently (Romance language) had intermediate performance, and the language group that does not have gender markings (English) had the worst performance. The authors concluded that transfer from their L1 gender systems helped the participants in processing the L2 gender markings, keeping in mind that these were all very proficient L2 Dutch speakers, so their performance was very high overall. These results provide evidence for transfer-based models because the most similar gender system proved to be the most helpful when processing L2.

Later, Sabourin and Stowe (2008) extended these findings using ERP; native Dutch speakers showed a P600 effect for two ungrammatical constructions, verbal domain dependency and grammatical gender agreement. Participants were shown the sentences one word at a time while recording ERPs from the scalp. Participants made grammaticality judgments about the sentences. For the syntactic construction, verbal domain dependency, German-Dutch and Romance language-Dutch bilinguals showed native-like processing (i.e. all groups showed a P600 to the syntactic violation); the construction was similar in the L1s of both groups, even though the Romance-language group has a different surface word order. However, German-Dutch and Romance language-Dutch bilinguals showed differential results for the morphosyntactic violation, grammatical gender agreements, such that a P600 was seen only if L1 and L2 were similar in that construction. The authors draw a similar conclusion to the one mentioned earlier: similarly governed processing routines in L1 and L2 have shared neural processing and transfer from L1 to L2 can best explain the pattern of results. With similar constructions, transfer succeeds because the processing routines are similar across the different languages. But constructions that are instantiated differently have different processing routines and therefore, transfer fails because there is not enough congruence between the structures. In this case, competition takes over and processing is not mirror native-like.

Chen, Shu, Liu, Zhao, and Li (2007) used ERP to examine subject-verb agreement in L2 learning. They examined how Mandarin-English bilinguals process subject-verb agreement in English online (ERP) and offline (grammatical judgment task) because Mandarin does not have subject-verb agreement. Despite high offline accuracy (88%), the Mandarin-English bilinguals did not show any online sensitivity (in the P600 component) to ungrammatical sentences in English for this unique construction; native English speakers do show such sensitivity. Chen et al. attributed this differential processing to L1 transfer because native Mandarin speakers do not have subject-verb agreement, so during online processing they show no sensitivity to the violation. The authors concluded that cross-linguistic similarity played an important role in the results of the study because the native English speakers and the native Mandarin speakers showed differences in the online measures.

Tokowicz and Warren (2010) extended these results to the self-paced reading paradigm, which allowed participants to control how long they looked at a word, in contrast to ERP studies, which typically have a set presentation time for each word. They analyzed online (self-paced reading) and offline (grammaticality judgment task) sensitivity to different, unique, and similar morphosyntactic structures in native English speakers taking introductory Spanish courses. They analyzed the word at which the violation occurred (target), the word following the target word (n+1) for spill-over effects, and the sentence-final word for wrap-up effects. Tokowicz and Warren replicated Tokowicz and MacWhinney's (2005) online sensitivity to similar constructions, but did not find online sensitivity in the unique condition until word n+1, finding it instead in the different condition at the target word. They attributed this discrepancy to a difference in stimuli and certain grammatical cues being stronger than others. Interestingly, one of the constructions in the similar condition (demonstrative adjective noun number agreement) seemed to be processed online in a more native-like way than the other L2 Spanish constructions they tested, but did not show this difference in the offline judgments. For example, ungrammatical sentences were responded to more accurately when more words that followed the target word agreed with it in number and/or gender. Therefore, the additional information after the target word provided more cues to the participants regarding grammaticality. These results do not completely confirm UCM predictions because participants showed online sensitivity to violations regardless of whether the morphosyntax was different or similar in L1 and L2, contrary to the predictions that sensitivity should be higher in the similar condition than in the different condition.

1.1 CURRENT STUDY

The present study tests the predictions of the UCM that syntactic similarities and differences affect L2 processing. Specifically, this experiment examines how sensitive L2 learners are to ungrammatical stimuli as a function of whether they occur in constructions that are cross-linguistically similar (Arabic and Mandarin), different (Arabic), or unique (Mandarin), as detailed below. The role of cue strength (availability and reliability) is also examined in this study and how it influences L2 processing in non-native speakers. The ungrammatical sentences have violations that vary cross-linguistically across the languages, but they also have different cue strengths in the sentential violations. As detailed above, several studies have examined morphosyntactic constructions using cross-language similarity as a basis for judgment using grammaticality judgments, self-paced reading, and ERPs (Chen et al., 2007; Sabourin et al., 2006; Sabourin & Stowe, 2008), but previous studies examining purely syntactic structures have focused almost exclusively on relative clause attachment preferences (see Dussias, 2003; Dussias & Sagarra, 2007; Frenck-Mestre & Pynte, 1997; Witzel et al., 2009). This exclusivity has led to a closed focus on the aspects of L2 syntactic processing that are involved in relative clause attachment. Looking at different types of syntactic processing, such as the basic syntactic effects of word order cues, can help expand the body of work on L2 processing in this area and inform us on different ways L2 speakers proceed during L2 comprehension other than relative clause attachments and morphosyntactic agreement. This study extends previous findings on relative clause attachment preferences (Dussias, 2003) to other syntactic constructions, consisting of pure word order violations that vary in their similarity to the participants' L1 and also in the strength of their cues.

In this experiment, we used eye-tracking to test the hypotheses about cross-language similarity based on the UCM and SSH. Previous studies have used behavioral methods such as self-paced reading (e.g., Tokowicz & Warren, 2010), which allows participants to control the amount of time they spend reading each word as they move across the sentence. In this way, self-paced reading captures processing difficulties by comparing reading times on certain words or areas that contain violations to those same areas without violations. Similarly, eye-tracking provides reading times for different words or regions in a sentence, but is able to capture more fine-grained information during reading, such as first and second pass reading times and regressions made in or out of specific regions. The underlying assumption is that covert cognitive attention and processing can be reflected through overt eye movements, which are considered to be almost automatic when reading (e.g., Just, Carpenter, & Woolley, 1982; Rayner, 2009). Eye-tracking takes advantage of the detailed information that can be extracted during the course of reading a sentence.

In particular, native Arabic and native Mandarin learners of English were tested with two types of syntactic violations. In an ungrammatical noun-article condition, the article followed the noun. In both Arabic and English, the definite article (‘the’ and ‘al’) comes before the noun. Thus, this was considered a “similar” construction for Arabic-English bilinguals. However, in Mandarin, there is no true definite article. Instead, Mandarin precedes its nouns with indefinite articles and articles that enumerate the noun. Thus, this was considered a “unique” construction for the native Mandarin speakers. However, because Mandarin does employ other types of articles, it may follow that this condition will be processed as a more “similar” structure.

In an ungrammatical noun-adjective condition, the adjective appeared after the noun. For adjectives, unlike English, Arabic places the adjective after the noun, so this was considered a “different” construction for Arabic-English bilinguals. However, in Mandarin, as in English, Mandarin places the adjective before nouns. Thus, this was considered a “similar” construction for the Mandarin-English bilinguals. To summarize, the native Arabic speakers were presented with both similar and different constructions and the native Mandarin speakers were presented with similar and unique constructions.

Note that the definite article construction seems to provide a stronger syntactic cue than the adjective condition, because definite articles come before their subsequent nouns more reliably than adjectives come before a noun. In other words, seeing a definite article *after* a noun (without that noun having an article in front of it) could be more reliably incorrect than having an adjective occur after a noun. For example, reading “*Tom sat on couch the…*” is more reliably incorrect than “*Tom sat on the couch comfortable…*”. Although the latter is ungrammatical, it is possible that the writer forgot a comma and meant to say something such as “*Tom sat on the couch, comfortable with pillows…*”. For our hypotheses, this interpretation of the noun-article condition suggests that native Mandarin speakers may be able to process this construction as a similar construction or will show a reliable effect because of the strength of the syntactic cue. In this same vein, we may see the same effect of this highly reliable syntactic cue for native English and native Arabic speakers such that it would override certain similarity effects. This argument will be addressed more in detail in the discussion.

We used several measures of sensitivity: accuracy and response times from a grammaticality judgment task performed at the end of each sentence, and reading time measures at the target region (article, adjective, and noun) and the post-target region (two word region after each critical region). We used standard measures taken during eye-tracking experiments that reflect the processing during reading (Warren, McConnell, & Rayner, 2008), including: first fixation durations, first pass reading times, first pass regressions out, go-past times, and total times spent on the target and post-target region. First fixation duration measures the amount of time a region is first fixated, which is the amount of time the subjects spend on the article, adjective, and noun combination the first time they fixate it. First pass reading time sums all the fixations on a region from the time the eye enters that region until it leaves, so if a region is fixated only once, it is equal to the first fixation duration. However, if a region is fixated more than once in reading across from left to right, first pass reading time will include these times. First pass regressions out of a region measures the percentage of trials that a region is exited during first pass reading with a regressive eye movement. Go-past times measure the amount of time spent on a region from entering that region to leaving it, so it may equal first pass reading time, but if the region had a regression from within it, then go-past times will also include fixations across previously seen material and any refixations on the region before it is exited. Finally, total time measures the total time spent on the region, including first and second pass readings of the region, so it will include rereading of the region and regressions made back into it from either the right or left (Warren, 2011). These will provide information on the extent to which L2 learners are sensitive to anomalies in English sentences.

First fixation durations did not provide discriminating information because parafoveal preview could have allowed participants to immediately identify the ungrammatical noun-article condition as ungrammatical because it had no article for each noun. For example, a sentence that would be ungrammatical on the first word of the critical region might read "*He used *spoon the clean for ice cream*". The noun-adjective and grammatical conditions both began the target region with "*the*" instead of "*spoon*" so it is not ungrammatical at the first word the participant fixates. The ungrammatical noun-adjective condition was only ungrammatical once the participant reached the third word, the adjective, in the region. Therefore, first fixation durations did not provide a good test of the differences between these conditions. Thus, the remaining measures were examined more closely in the analyses.

Longer reading times for ungrammatical sentences compared to the grammatical sentences will be taken to indicate sensitivity to a particular grammatical violation; we expect that native English speakers will be sensitive to both types of violations. If the results conform to UCM predictions, L2 speakers will show more sensitivity to the syntactic violations that are similar to L1 through longer reading times (more disruption during reading) for these sentences. Violations that are unique to L2 may result in sensitivity to the structures depending on cue reliability and availability. Specifically, the native Arabic speakers will be less sensitive to violations in the ungrammatical different (noun adjective) condition, but will be more sensitive to violations in the ungrammatical similar (noun article) condition. This pattern is expected because the noun-adjective condition will provide competition for the native Arabic speakers –

the word order structure is instantiated differently across languages and is therefore more difficult to process in a native-like way. The similar condition will show positive transfer from L1 because these structures are similar in English and Arabic and it will be easier to process these in a more native-like way. The native Mandarin speakers should show sensitivity to the similar (noun adjective) condition because there will be positive transfer from the processing system for the violation, which will allow them to process these in a more native-like way. The unique (noun article) construction provides neither transfer nor competition because it is not present in the L1. Therefore, there may be sensitivity to violations in this condition, but it will depend on cue strength. Cross-language similarity and cue strength are two facets addressed by the UCM and these two necessarily interact during L2 processing. Because there is no true definite article in Mandarin, these speakers may not attend to the ungrammaticality at all, showing similar results to those of Chen et al. (2007) in which native Mandarin speakers showed no differences in online processing for a construction unique to L2. On the other hand, because Mandarin does use other types of articles and the noun-article violation represents a stronger cue than the noun-adjective condition, Mandarin speakers may show increased sensitivity to this construction and the noun-article condition may produce results comparable to the similar condition. In this way, the similarity conditions and cue strength in each sentence may provide different results with respect to how L2 speakers process these violations.

Because these are structures that require higher-level processing (they are at the syntactic level, not the word level), the SSH would predict that there should be no transfer from the L1 to the L2. Therefore, differences between the similarity conditions should not arise. This model would also claim that the non-native speaker groups would not show native-like processing because their L2 grammar representations are too shallow to achieve this. The native Arabic and native Mandarin speakers should not show sensitivity to the English violations because they have no transfer from their L1 and they should not show native-like processing when compared to the native English control group. However, Clahsen and Felser (2006b) also cite evidence that late L2 learners can achieve native-like processing in grammatical relationships that involve local dependencies such as subject-verb or gender agreement. Therefore, because all of the violations are local syntactic dependencies in this study, the L2 speakers may show native-like sensitivity to the violations, but there should not be modulation based on the similarity condition of each native language group. Additionally, our results may differ based on the syntactic cue strength of the constructions, as mentioned above. The difference in cue strength between the noun-adjective and noun-article condition may lead to differences in online processing that do not conform to the UCM similarity predictions alone. The UCM addresses both transfer and cue strength effects in L2 processing, and therefore, the two types of information necessarily interact. The effects of a more reliable cue over a less reliable cue would show basic syntactic effects of word order in L2 speakers as well as native English speakers, and this could show native-like processing with a strong enough cue. The SSH does not address cue strength like the UCM does, but does allow that native-like processing of basic word order (a local dependency) may be learned. Examining how cue strength interacts with similarity will provide more information and evidence for both the UCM and the SSH.

2.0 METHODS & RESULTS

2.1 METHODS

2.1.1 Design.

We used a three-language group (native English, native Arabic, native Mandarin speakers) x 3 grammaticality (grammatical, ungrammatical noun-adjective, ungrammatical noun-article) within-participants design.

2.1.2 Participants.

A total of 29 native Arabic speakers, 31 native Mandarin speakers, and 29 native English speakers participated in this study. To perfectly match the counterbalancing of items, participants were removed from the analyses to ensure an equal number per list version and per group. This resulted in the removal of nine native English participants, nine native Arabic participants, and eleven native Mandarin participants for a total of 20 native English speakers, 20 native Arabic speakers and 20 native Mandarin speakers.

All non-native English speakers were recruited from the English Language Institute (ELI) at the University of Pittsburgh, using in-class announcements, advertisements, and recruitment posters. Participants received $20 as compensation for their participation in the study. The native

English speakers were students in an introductory psychology course at the University of Pittsburgh and received course credit for their participation.

The native Arabic speakers were current students at the ELI, ranging from levels 3 to 5 (out of 5 levels), from intermediate to advanced in four classes: reading, writing, listening and speaking. This corresponded to a Michigan Test of Language Proficiency (MTELP) score of 55 or above (out of 100). The majority of the native Mandarin speakers were also current students at the ELI, with the exception that eight were students in the Department of Linguistics at the University of Pittsburgh. These students were taking English language classes in their department similar to those offered at the ELI in preparation to become undergraduates at the University of Pittsburgh. They were taking these language courses because their level of English proficiency was not sufficient to enable them to begin taking courses instructed in English. The students were enrolled in reading, writing and speaking classes, much like the ELI classes. The native English speakers were recruited from the University of Pittsburgh's Psychology subject pool, and were given course credit for their participation in the study.

All participants completed a questionnaire detailing their language experiences (Tokowicz, Michael, & Kroll, 2004). These measures included reading proficiency, writing proficiency, conversational fluency, and spoken language comprehension. The ratings indicate that even though all three language groups had a very high self-rated ability across all measures in L1, their L2 measures were indicative of less ability in their L2s, demonstrating a clear dominant language for all three language groups (see Table 1 for detailed comparisons between groups).

Table 1. *Means for self-rated proficiency across all three language groups. Standard deviations are shown in parentheses.*

	Native English speakers		**Native Arabic speakers**		**Native Mandarin speakers**	
Gender [a]	M	F	M	F	M	F
N	12	7	15	4	6	14
Age (years)	18.80 (1.77)		25.20 (3.43)		25.90 (5.64)	
Time in USA (months)	223.80 (23.45)		8.45 (5.66)		6.63 (6.59)	
			Self-rated proficiency [b]			
L1 Reading	9.35 (0.88)		9.65 (0.67)		9.40 (0.75)	
L1 Writing	9.30 (1.03)		9.00 (1.12)		8.65 (1.14)	
L1 Speaking	9.65 (0.75)		9.75 (0.55)		9.80 (0.52)	
L1 Listening	9.75 (0.55)		9.80 (0.41)		9.55 (0.69)	
L2 [c] Reading	4.20 (2.46)*		6.30 (1.34)		5.28 (1.67)	
L2 Writing	4.05 (2.67)*		5.70 (1.22)		4.60 (1.57)	
L2 Speaking	3.30 (2.25)*		6.45 (1.96)		5.50 (1.47)	
L2 Listening	4.05 (2.37)*		6.65 (1.66)		5.00 (1.17)	

[a] One native Arabic participant and one native English participant did not include their gender. [b] Self-rated proficiency is on a 1 to 10 point Likert-type scale with 10 indicating the highest level of ability and 1 indicating the lowest level of ability. [c] L2s for native English speakers included Spanish (n=4), French (n=4), German (n=1), and none (11).

2.1.3 Stimuli.

The stimuli consisted of 56 experimental English sentences and 4 practice items. These 56 sentences were divided into 28 grammatical sentences, 14 ungrammatical sentences with a noun-adjective violation (the adjective appeared after the noun), and 14 ungrammatical sentences with a noun-article violation (the article appeared after the noun).

The target region in each sentence was the group of three words that were manipulated to achieve the syntactic violation. Therefore, the target region included the same three words across the sentence types. This was done so that eye-tracking measures were always taken from the same three words, albeit in a different order. Participants were exposed to only one of the three versions of each sentence; sentences were counterbalanced across participants. See examples below; the bolded words indicate the target region.

EXAMPLES:	Grammatical:	*She demanded **the pink cake** for her birthday.*
	Ungram. Noun-Adj.:	*She demanded **the cake pink** for her birthday.*
	Ungram. Noun-Art.:	*She demanded **cake the pink** for her birthday.*

Experimental sentences were divided into five interest areas. For example, the sentence "She demanded the pink cake for her birthday" was divided into the following areas: She/demanded/the pink cake/for her/birthday. The target region was the third interest area.

Sentences were counterbalanced across participants such that each participant saw each sentence only once, in the grammatical form, the ungrammatical noun-adjective form, or the ungrammatical noun-article form. This resulted in four list versions of the stimuli, to which participants were randomly assigned. There were twice as many grammatical as ungrammatical sentences so four list versions were needed to ensure that participants saw an equal number of grammatical and ungrammatical sentences. The words were also crosschecked with a list of words with which students in the ELI at the University of Pittsburgh should be familiar such that the majority of the words used in the stimuli were on the list. This included an updated version of the General Service List (Bauman, 2010; West, 1953) and all vocabulary included in or assumed to already be known in the first four chapters of the ELI vocabulary textbooks, *Words for*

Students of English (Rogerson, Davis, Hershelman, & Jasnow, 1992; Rogerson, Esarey, Schmandt, & Smith, 1992; Rogerson, Hershelman, Jasnow, & Moltz, 1992).

Fifty-six filler sentences were also presented, which consisted of 28 grammatical and 28 ungrammatical sentences involving morphosyntactic violations. These sentences were taken from Tokowicz and Warren (2010); one quarter included violations of subject verb person agreement and one quarter included violations of verb aspect licensing.

2.1.4 Procedure.

Participants were instructed that after each sentence, they would see a question asking them if the sentence they saw was grammatical, and they were to respond yes or no by pressing a button. The apparatus was an EyeLink 1000 tower-mounted eye-tracker (SR Research Ltd., Toronto, Ontario, Canada). The average eye gaze position accuracy ranged from 0.05 to 0.25 visual degrees. Data were recorded monocularly from the pupil of the right eye at a sampling rate of 1000Hz. Chin and forehead rests were used to minimize head movement. The screen resolution was set at 1024 x 768 pixels and the stimuli were presented in black Times New Roman 20-point font on a white background. Sentences were left justified. Standard procedures for an eye-tracking experiment were observed (Warren, McConnell & Rayner, 2008). Before beginning the experiment, the eye-tracker was calibrated to each participant using a nine-point calibration grid, followed by a validation check. Before each sentence, a one-point calibration check shown on the left side of the screen was conducted to ensure that participants began reading the sentence at the leftmost point consistently.

After the one-point calibration check, participants pushed a button while looking at this point to begin reading a sentence. The sentence was then presented until the participant pressed the button again for the question. After each question, the participant answered yes or no by pressing one of two buttons on the controller. After the response, the one-point calibration check screen reappeared, at which point the participant could move on to the next trial by pushing the button. The eye-tracking task took on average 30 minutes to complete, and participants then completed a vocabulary post-test and the language history questionnaire.

2.2 RESULTS

2.2.1 Data trimming.

Eye-tracking data were cleaned following standard procedures (Warren et al., 2008; Warren, 2011), according to which single fixations shorter than 80 ms that fell within 1.5 characters of another fixation were combined into the longer fixation. Fixations longer than 1000 ms were eliminated. Skipped trials were also removed from the analyses; for the target region, this resulted in the removal of 1.97% of the data for the native Arabic-speaking participants, 0.70% of the data for the native English-speaking participants, and 0.37% of the data for the native Mandarin-speaking participants.

2.2.2 Eye-tracking measures.

Five main measures from the eye-tracking data were computed: first fixation durations, first pass reading times, first pass regressions out, go-past time, and total time. Each measure was analyzed using a 3 native language (native English, native Arabic, native Mandarin) by 3 grammaticality condition (grammatical, ungrammatical noun-adjective, ungrammatical noun-article) repeated measures ANOVA. For analyses violating the assumption of sphericity, the Greenhouse-Geisser correction was applied. Following convention (Picton et al., 2000), we report uncorrected degrees of freedom, the corrected *p*-value, and the corrected mean square error values. For all post-hoc analyses performed, the Bonferroni correction was applied, resulting in a more conservative alpha level needed to reject the null hypothesis. Of interest to us is the comparison between the grammatical sentences to the two ungrammatical sentences (noun-adjective and noun-article), therefore all analyses address these comparisons. We first present the interactions by measure, because these provide the strongest tests of the predictions. We then report significant main effects of native language and grammaticality.

2.2.3 Interactions in the target region.

Detailed below are grammaticality by native language interactions in the five eye-tracking measures. These are of primary interest for this study because we are interested in whether sensitivity to grammaticality differs as a function of native language. See Table 2 for a summary of reading time measures across language groups for the target region.

Table 2. *Grammaticality vs. ungrammaticality effects in target region and reading time measures in the three language groups in mean milliseconds (ms).*

	English			Arabic			Mandarin		
Measure	*Gramm.*	*Noun-Adj.*	*Noun-Art.*	*Gramm.*	*Noun-Adj.*	*Noun-Art.*	*Gramm.*	*Noun-Adj.*	*Noun-Art.*
First fixation duration (ms)	218_{a}	$219_{a,b}$	242_{b}	272_{a}	264_{a}	310_{b}	230_{a}	229_{a}	264_{b}
First pass reading time (ms)	544_{a}	681_{b}	757_{b}	1401_{a}	1517_{a}	1371_{a}	1064_{a}	1042_{a}	1042_{a}
Regressions out[a]	0.16_{a}	0.17_{a}	0.36_{b}	0.26_{a}	0.28_{a}	0.40_{b}	0.13_{a}	$0.23_{a,b}$	0.33_{b}
Go past time (ms)	680_{a}	818_{b}	1127_{c}	1914_{a}	2242_{b}	2518_{b}	558_{a}	597_{a}	734_{a}
Total time (ms)	917_{a}	1179_{b}	1282_{b}	3503_{a}	4132_{a}	3830_{a}	2385_{a}	3100_{b}	3296_{b}

Note: Means in the same row that do not share an alphabetic subscript differ at the Bonferroni-corrected $p < .008$ level. [a] "Regressions out" are measured in average percent of regressions exiting the region, not ms.

2.2.3.1 *First fixation durations*.

There was no significant interaction between grammaticality and native language group, $F < 1$ (see Figure 1).

2.2.3.2 *First pass reading times*.

Grammaticality interacted with native language group, F (4, 114) = 3.92, MSE = 131771.06, $p < .05$, such that the native English speaking participants showed significantly longer reading times for both types of ungrammatical sentences than for the grammatical sentences (see Figure 2). By contrast, both non-native language groups showed no significant difference between ungrammatical and grammatical sentences.

2.2.3.3 ***Regressions out*****.**

There was no significant interaction between grammaticality and native language group, *F* (2,114) = 1.30, *MSE* = .022, *p* = .28 (see Figure 3).

2.2.3.4 ***Go-past times*****.**

There was an interaction between grammaticality and native language group, *F* (4, 114) = 3.06, *MSE* = 353218.38, $p < .05$ (see Figure 4). Post-hoc analyses showed that native Arabic speakers showed significantly longer go-past times for both the noun-article, *t* (59) = 3.49, $p < .01$, and the noun-adjective conditions, *t* (59) = 3.48, $p < .01$, than the grammatical condition; the differences between the two ungrammatical conditions was marginally significant, due to the Bonferroni correction, *t*(59) = 2.13, *p* = .04. The native English speakers demonstrated similar results to the native Arabic speakers: they spent significantly less time on the grammatical condition than the noun-article condition, *t* (19) = 7.48, $p < .001$, and the noun-adjective condition, *t* (19) = 4.04, $p < .01$. There was also a significant difference between the two ungrammatical conditions for the native English speakers, *t* (19) = 5.68, $p < .001$. For native Mandarin speakers, the ungrammatical noun-article condition had marginally longer go-past times than the grammatical condition after correction, *t* (59) = 2.36, *p* = .03, and marginally longer go-past times than the noun-adjective condition, *t* (59) = 2.24, *p* = . 03). There was no significant difference between the grammatical and the noun-adjective conditions.

2.2.3.5 ***Total time***.

There was a marginal interaction between native language group and grammaticality, *F* (4,114) = 2.40, MSE = 781597.98, *p* = .054 (see Figure 5). Corrected post-hoc t-tests demonstrated that native English speakers showed longer total times for the noun-article condition than the grammatical condition, *t* (19) = 4.28, *p* < .001. The same pattern emerged for the noun-adjective condition compared to the grammatical condition, *t* (19) = 4.27, *p* < .001. Native Mandarin speakers showed the same effect with longer total time spent on the noun-article condition than the grammatical condition, *t* (19) = 4.88, *p* < .001, and the noun-adjective condition, *t* (19) = 3.64, *p* < .01. However, native Arabic speakers had marginally longer total times for the ungrammatical noun-adjective sentences than grammatical sentences (*t* (19) = 2.32, *p* = .03), but there was no difference between the grammatical sentences and the ungrammatical noun-article sentences.

Figure 1. First fixation durations for the target region

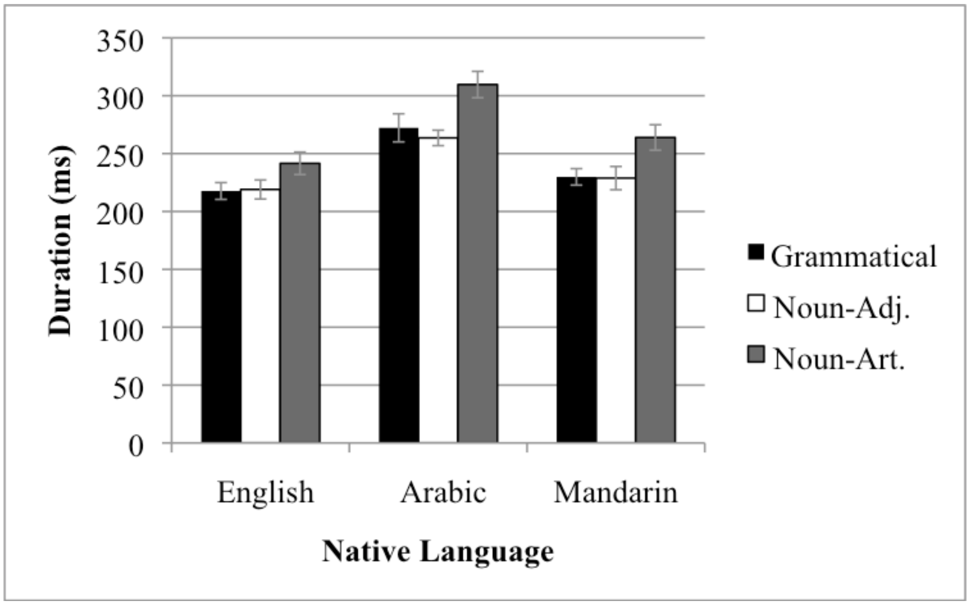


Figure 2. First pass reading times for the target region

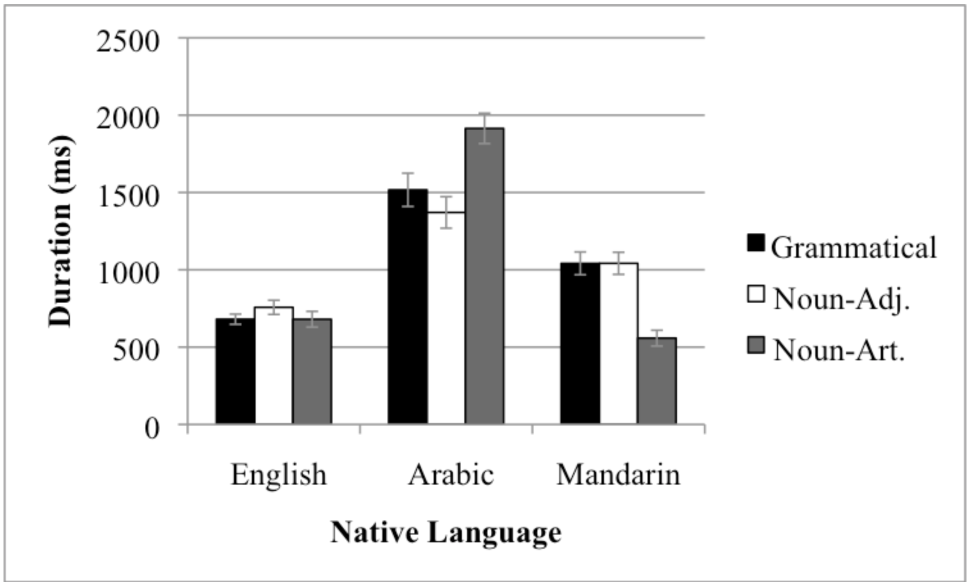


Figure 3. Regressions out for the target region

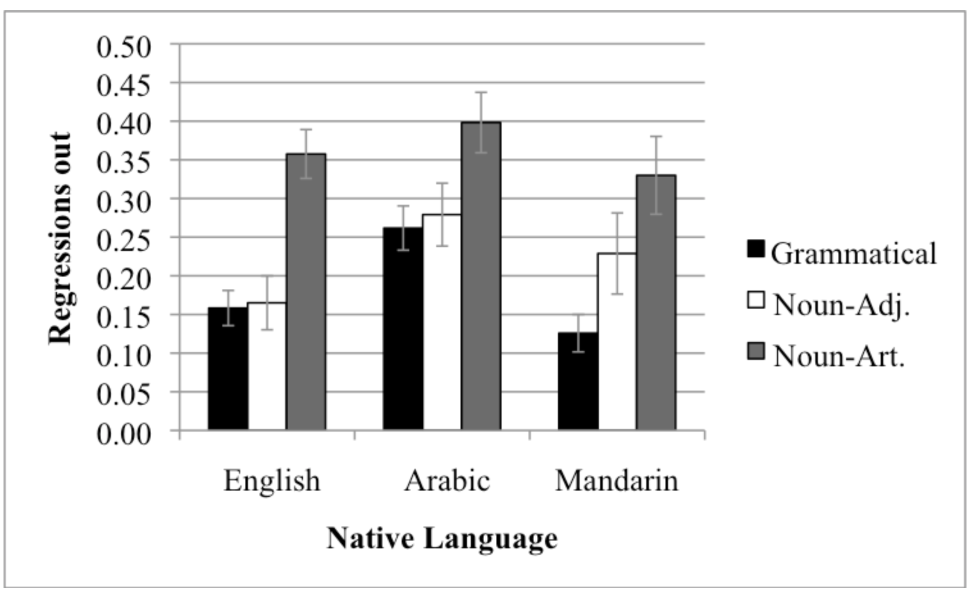


Figure 4. Go past times for the target region

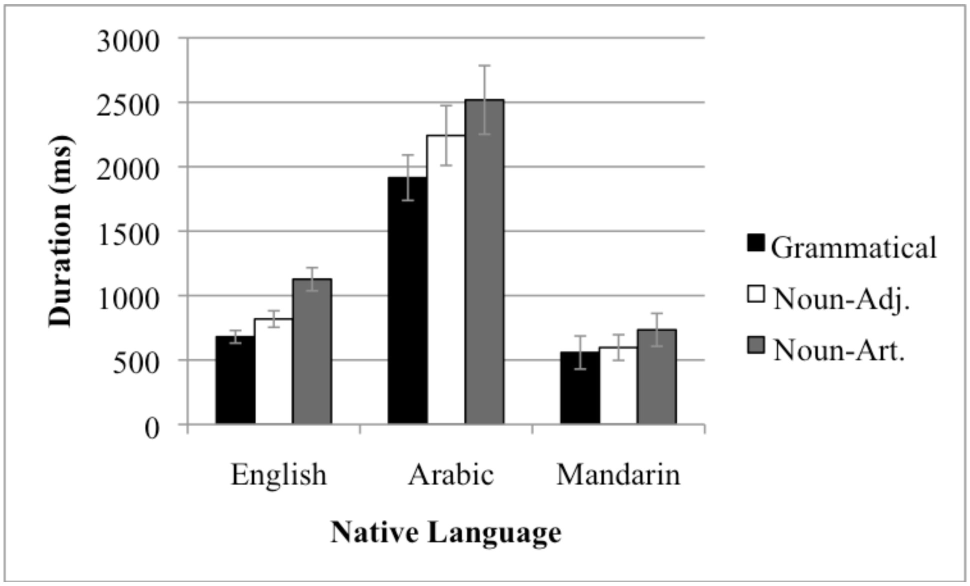
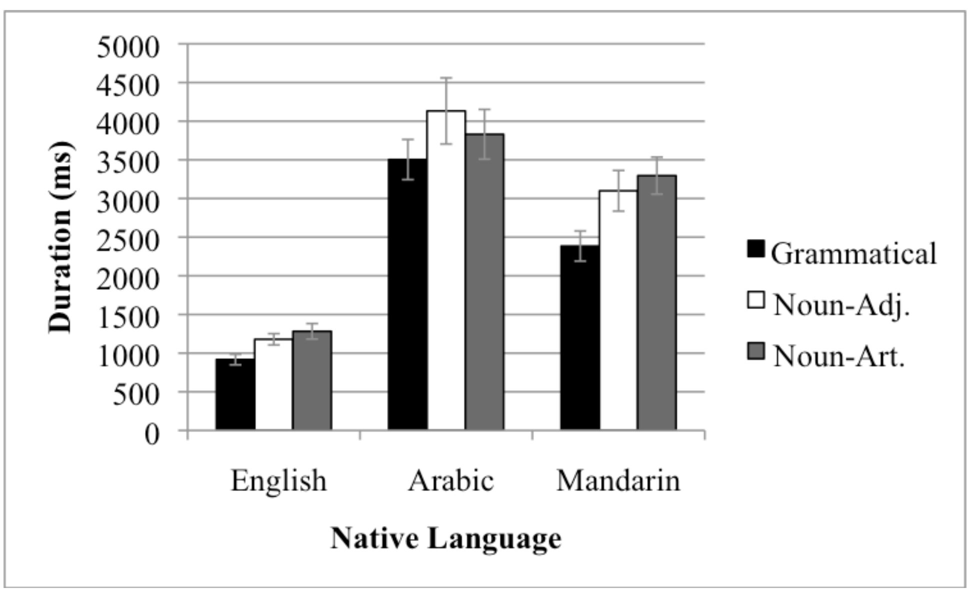


Figure 5. Total time for the target region



2.2.4 Interactions in the post-target region.

Detailed below are interaction effects between native language and grammaticality in the five eye-tracking measures for the post-target region. Previous research has shown that L2 speakers may have later effects during L2 processing because of differences in proficiency compared to native speakesr (Tokowicz & Warren, 2010). Therefore, post-target processing was measured to determine to see if there were any effects of spill-over processing after the target area.

2.2.4.1 *First fixation durations*.

There was no interaction between grammaticality and native language group (*Fs* > 1) for first fixation durations (see Figure 6).

2.2.4.2 *First pass reading times*.

There was no interaction between grammaticality and native language, F (4,114) = 1.24, MSE = 8965.80, p = 0.30 (see Figure 7).

2.2.4.3 *Regressions out*.

There was also a marginally-significant interaction, F (4, 114) = 2.21, MSE = .07, p = .07 (see Figure 8). Post-hoc analyses revealed that native Arabic speakers had significantly fewer regressions out in the grammatical condition than in the noun-adjective condition, t (19) = 5.44, p < .001. The noun-adjective condition did not differ significantly from the noun-article condition. Native Mandarin speakers had significantly fewer regressions out of the grammatical condition than the noun-adjective condition, t (19) = 3.48, $p < .01$, and marginally fewer regressions out than the noun-article condition, t (19) = 2.32, $p = .03$, and these two were not different from each other. Finally, the native English speakers showed a marginal difference only between the grammatical and noun-adjective sentences t (19) = 2.25, $p = .03$.

2.2.4.4 *Go-past times*.

There was a significant interaction between grammaticality and native language group, F (4,114) = 2.89, MSE = 322201.77, $p < .05$ (see Figure 9); post-hoc t-tests showed that native Mandarin speakers showed significantly longer go-past times for the noun-article condition, t (19) = 3.21, $p < .01$, and for the noun-adjective condition than the grammatical condition, t (19) = 3.78, $p < .001$. Native Arabic speakers showed a marginal difference between the grammatical and the noun-adjective sentences, t (19) = 2.85, $p = .01$, but the noun-adjective condition was not significantly different from the noun-article condition. Finally, native English speakers showed a significant difference only between the two ungrammatical sentences, t (19) = 2.35, $p = .03$.

2.2.4.5 *Total time*.

There was no significant interaction between grammaticality and native language group, F (4,114) = 1.18, MSE = 136838.30, $p = .32$ (see Figure 10).

Figure 6. First fixation durations for the target region

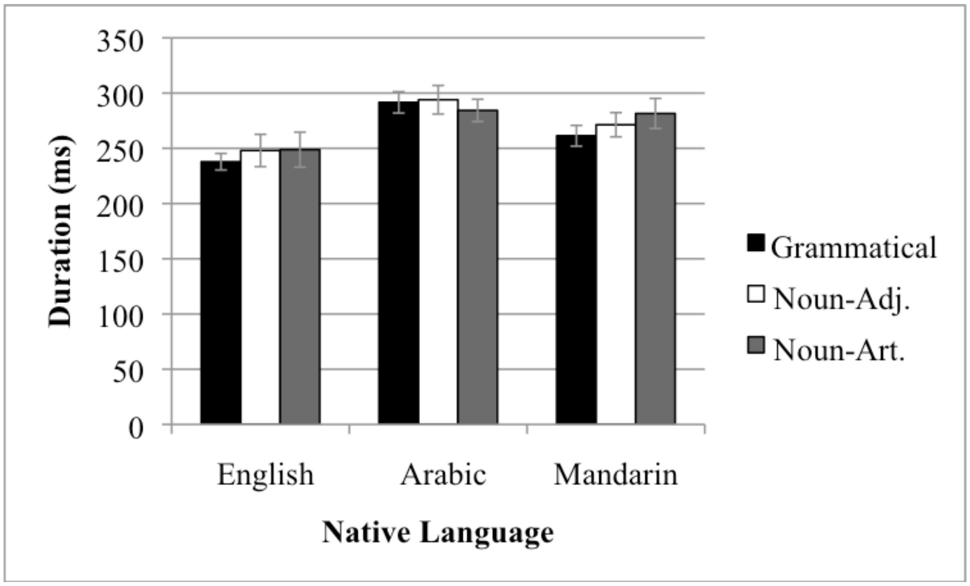


Figure 7. First pass reading times for the target region

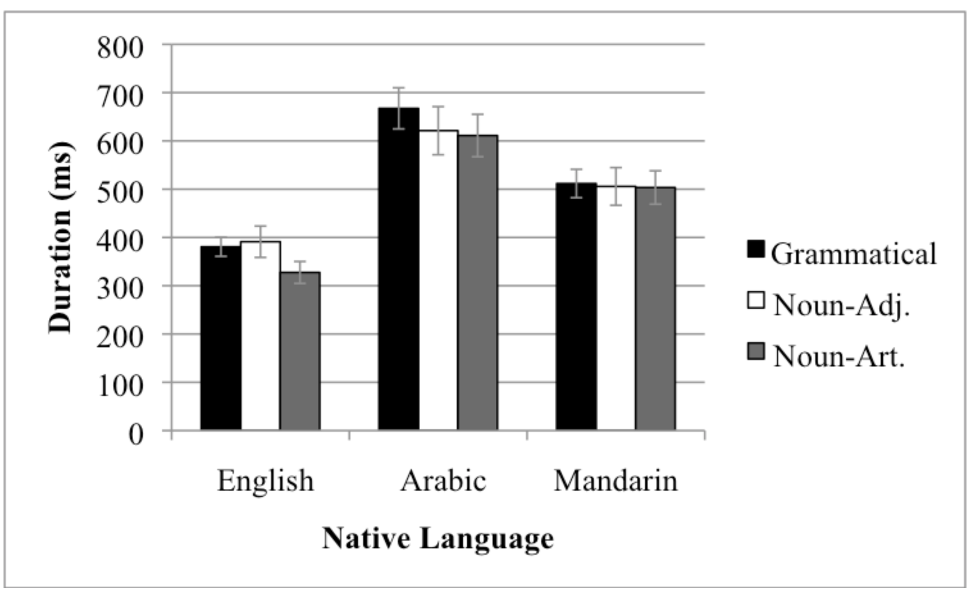


Figure 8. Regressions out for the target region

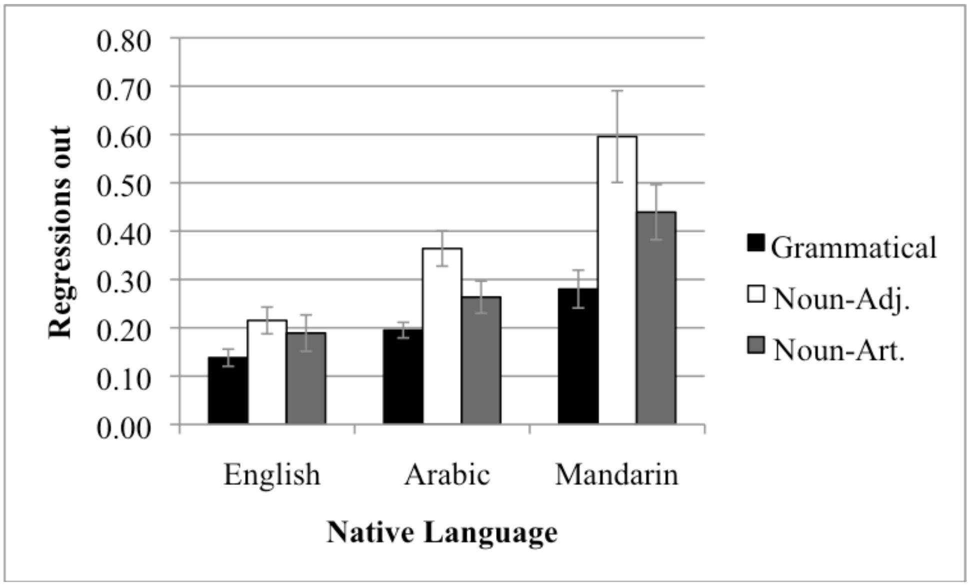


Figure 9. Go past times for the target region

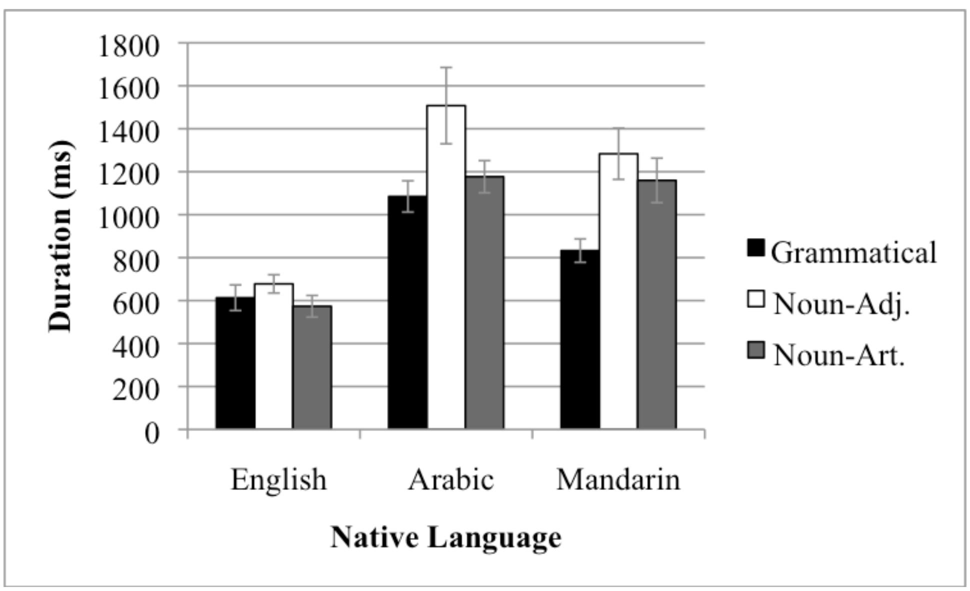
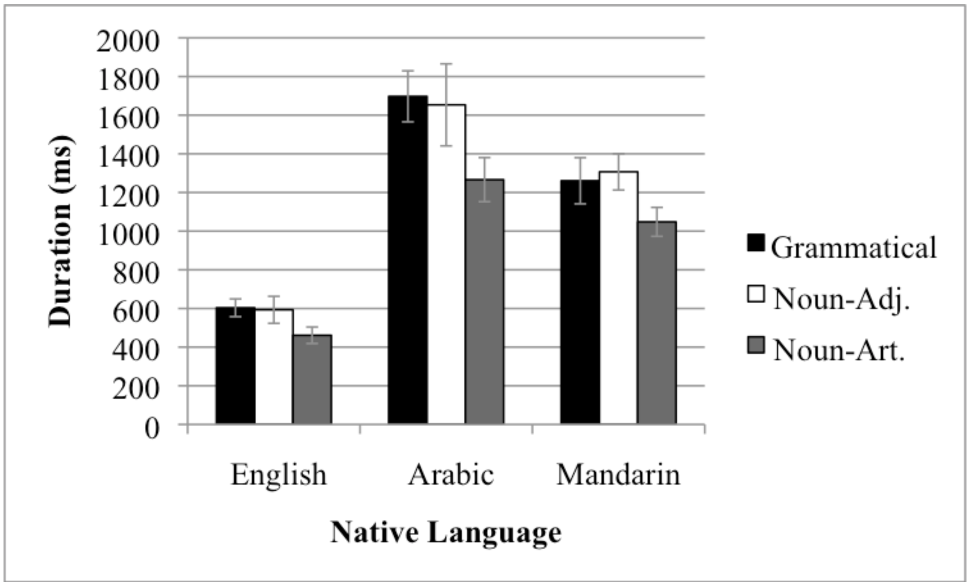


Figure 10. Total times for the target region



2.2.5 Main effects of native language in the target region.

We found main effects of native language, such that, generally, native English speakers had shorter looking times than native Arabic and native Mandarin speakers in the five measures. This was expected because the native Arabic and native Mandarin speakers were reading in L2. Detailed below are the native language effects across the five eye-tracking measures. See Table 3 for differences in magnitudes across reading time measures and language groups in the target region.

Table 3. *Magnitudes of effect for target region across reading time measures (ungrammatical conditions minus grammatical condition).*

	English		**Arabic**		**Mandarin**	
Measure	*Noun-Adj. – Gramm.*	*Noun-Art. – Gramm.*	*Noun-Adj. – Gramm.*	*Noun-Art. – Gramm.*	*Noun-Adj. – Gramm.*	*Noun-Art. – Gramm.*
First fixation duration (ms)	26	34	28	49	30	42
First pass reading time (ms)	152	228	193	261	153	205
Regressions out	0.11*	0.20	0.12	0.17	0.16	0.22
Go past time (ms)	172*	447	361*	643	151*	271
Total time (ms)	298‡	392	874	731	891	1024

2.2.5.1 ***First fixation durations*****.**

There was a significant effect of native language group, *F* (2,57) = 15.03, *MSE* = 50017.04, $p < .001$; corrected post-hoc tests showed that native Arabic speakers had significantly longer first fixation durations than native English participants (282 v. 226 ms) ($p < .001$) and native Mandarin participants (282 v. 241 ms) ($p < .01$). Native Mandarin participants did not have longer fixations than native English participants. See Figure 1 for means and standard errors.

2.2.5.2 ***First pass reading times*****.**

There was a significant effect of native language group, *F* (2,57) = 32.88, *MSE* = 8873698.10, $p < .001$; corrected post-hoc tests showed that the native Arabic speakers showed longer first pass reading times than the native Mandarin (1430 vs. 1049 ms) ($p < .01$) and native English speakers (1430 vs. 661 ms) ($p < .001$). Native Mandarin participants showed longer first pass reading times than native English participants ($p < .001$). See Figure 2 for means and standard errors.

2.2.5.3 ***Regressions out*.**

There was a marginal effect of native language group, F (2,57) = 2.90, MSE = .15, p = .06, such that the native Arabic speakers showed marginally more regressions out (0.31) than either the native Mandarin (0.23) or native English (0.23) speakers. See Figure 3 for means and standard errors.

2.2.5.4 ***Go past times*.**

There was an effect of native language group, F (2,57) = 35.32, MSE = 44268637.20, $p < .001$, such that the native Arabic speakers showed the longest go-past times overall (2225 ms) ($p < .001$), but there was no difference between the native English speakers (875 ms) and the native Mandarin speakers (630 ms) ($p = .71$). See Figure 4 for means and standard errors.

2.2.5.5 ***Total times*.**

There was a significant effect of native language, F (2,57) = 38.50, MSE = 113100000, $p < .001$; corrected post-hoc multiple comparisons showed that native Arabic participants spent more time on the region (3822 ms) than the native Mandarin speakers (2927 ms) ($p < .05$) and the native English speakers (1126 ms) ($p < .001$). Native English speakers spent the least amount of time on the target region ($p < .001$). See Figure 5 for means and standard errors.

2.2.6 Main effects of grammaticality in the target region.

Main effects of grammaticality were found, in all five eye-tracking measures. These main effects indicated that our grammaticality manipulations were effective. Detailed below are the grammaticality effects across the five eye-tracking measures.

2.2.6.1 *First fixation durations*.

See Figure 1 for means and standard errors for first fixation durations for each native language group in the target region. For first fixation durations in the target region, there was a significant effect of grammaticality, F (2, 114) = 21.46, MSE = 22158.69, $p < .001$. Post-hoc *t*-tests revealed that the ungrammatical noun-article condition (272 ms) was fixated for significantly longer than the ungrammatical noun-adjective (237 ms), t (59) = 5.46, $p < .001$, and grammatical conditions (240 ms), t (59) = 5.33, $p < .001$. There were no significant differences in the magnitudes of the effects for first fixation durations.

2.2.6.2 *First pass reading times*.

Figure 2 shows means and standard errors for first pass reading times for each native language group in the target region. First pass reading measures in the target region showed a marginal effect of grammaticality, F (2, 114) = 2.75, MSE = 92667.82, $p = .06$), and post-hoc tests showed that the grammatical condition (1003 ms) had longer first pass reading times than the ungrammatical noun-adjective condition (1080ms), t (59) = 3.01, $p < .001$, although it was not significantly longer than the noun-article condition. There were no significant differences in difference scores for the magnitudes of the grammaticality effects.

2.2.6.3 ***Regressions out*****.**

Regression out measures showed only a significant effect for grammaticality (see Figure 3), F (2,114) = 31.57, MSE = 0.53, p < .001, and corrected post-hoc t-tests showed that the ungrammatical noun-article condition had significantly more regressions out of the region (0.36) than the ungrammatical noun-adjective condition (0.22), t (59) = 5.39, p < .001, and the grammatical condition (0.18), t (59) = 8.11, p < .001. Difference scores analyses demonstrated that native English speakers showed a significant difference in regressions out, such that the magnitude of difference between the noun-article and grammatical conditions (0.20) was significantly greater than the difference between noun-adjective and grammatical conditions (0.11), t (19) = -2.55, p < .05. This demonstrated greater disruption to the noun-article condition than the noun-adjective condition because, on average, this region was exited with more regressions, indicating that more difficulty was experienced with those sentences. See Table 3 for difference scores.

2.2.6.4 ***Go past times*****.**

For means and standard errors for go-past measures for each native language group, see Figure 4. There was a significant effect of grammaticality, F (2, 114) = 30.32, MSE = 3504394.91, p < .001; post-hoc t-tests showed that the grammatical condition had the shortest go-past time (1051 ms) than the noun-adjective, t (59) = 4.22, p < .001, and the noun-article (t (59) = 5.94, p < .001) conditions. The ungrammatical noun-adjective condition had significantly shorter go-past times (1219 ms) than the ungrammatical noun-article condition (1460 ms), t (59) = 4.71, p < .001.

Difference score results for the target region showed that all three native language groups showed the same pattern for go-past times. Native Arabic speakers showed a significant difference in magnitude for their go-past times, t (19) = -2.30, $p < .05$, such that the magnitude of difference between the noun-article and grammatical conditions (642 ms) was significantly larger than the difference between the noun-adjective and grammatical conditions (361 ms). Therefore, they had significantly more disruption to noun-article condition than the noun-adjective condition by spending more time (282 ms) on the noun-article condition than the noun-adjective condition. The native Mandarin speakers showed the same pattern demonstrating that they had a larger magnitude of difference between the noun-article and grammatical conditions (271 ms) than the noun-adjective and grammatical conditions (151 ms), t (19) = -2.67, $p < .05$. They showed more disruption as well to the noun-article condition than the noun-adjective condition by spending 120 ms more on the noun-article condition, although they did not show significant differences in mean go-past times between the three conditions. Similarly, the native English speakers had significantly larger magnitude of difference between the noun-article and grammatical condition (447 ms) than between the noun-adjective and grammatical conditions (172 ms), t (19) = -5.51, $p < .001$. Therefore, they also had significantly more disruption to the noun-article condition, spending 274 ms more on this condition than the noun-adjective condition.

2.2.6.5 *Total times*.

See Figure 5 for means and standard errors for total time spent on the target region. There was a significant effect of grammaticality, F (2,142) = 17.56, MSE = 5719730.47, $p < .001$, such that, overall, the grammatical condition had a shorter total time (2268 ms) than the noun-article condition (2803 ms), t (59) = 5.55, $p < .001$, and the noun-adjective condition (2804 ms), t (59) = 4.67, $p < .001$.

The difference score results for total time showed that only native English speakers had a marginal effect for total time, *t* (19) = -2.04, *p* = .056, such that the magnitude of difference between the grammatical and the noun-article conditions (392 ms) was marginally larger than the difference between the grammatical and the noun-adjective conditions (298 ms). This mirrors the go-past results detailed above, demonstrating that native English speakers again had more disruption to the noun-article condition than the noun-adjective condition.

2.2.7 Main effects of native language in the post-target region.

Detailed below are the grammaticality effects across the five eye-tracking measures in the post-target region.

2.2.7.1 *First fixation durations*.

See Figure 6 for means and standard errors. There was a significant effect of native language group, F (2, 57) = 5.20, *MSE* = 30795.01, $p < .01$, such that the first fixation durations for native Arabic and native Mandarin speakers were not significantly different (290 ms and 271 ms respectively), but native Arabic speakers had significantly longer post-target first fixations than native English speakers (245 ms) ($p < .01$). There were no significant differences between native Mandarin and native English speakers' post-target first fixation durations.

2.2.7.2 *First pass reading times*.

There was a significant effect of native language, F (2,57) = 16.69, *MSE* = 1068591.80, $p < .001$, and corrected post-hoc multiple comparisons showed that the native English speakers had shorter first pass reading times (366 ms) compared to native Arabic (633 ms) ($p < .001$) and native Mandarin speakers (507 ms) ($p < .05$) (see Figure 7 for means and standard errors). There was a significant difference between the native Arabic and native Mandarin speakers, such that the native Arabic speakers had longer reading times than the native Mandarin speakers ($p < .05$).

2.2.7.3 ***Regressions out.***

There was a significant effect of native language, F (2,57) = 17.02, MSE = 1.02, $p < .001$, and corrected post-hoc multiple comparisons tests showed that native Mandarin speakers showed significantly more regressions out (0.44) than either the native Arabic (0.27) ($p < .01$) or native English speakers (0.18) ($p < .001$). There was no significant difference between native Arabic and native English speakers. See Figure 8 for means and standard errors.

2.2.7.4 ***Go past times.***

There was a significant effect of native language group, F (2,57) = 21.79, MSE = 6515502.35, $p < .001$, and corrected post-hoc multiple comparisons tests showed that there was no significant difference between the native Arabic and native Mandarin speakers' go-past times, but native Arabic speakers had significantly longer measures than native English speakers (1256 ms v. 621 ms, respectively) ($p < .001$) and native Mandarin speakers (1092 ms) ($p < .001$) had significantly longer go past-times than native English speakers. Refer to Figure 9 for means and standard errors.

2.2.7.5 ***Total times.***

There was an effect of native language group, F (2,57) = 29.01, MSE = 15101948.10, $p < .001$, and corrected post-hoc multiple comparisons revealed that the native Arabic speakers had longer total times (1539 ms) than the native English speakers (553 ms) ($p < .001$) and the native

Mandarin speakers (1205 ms) ($p < .05$). The difference between the native English and native Mandarin speakers was also significant ($p < .001$), such that the native Mandarin speakers had longer total times on the post-target region than the native English speakers. See Figure 10 for means and standard errors.

2.2.8 Main effects of grammaticality in the post-target region.

Detailed below are grammaticality effects for the post-target region across the five eye-tracking measures.

2.2.8.1 *First fixation durations*.

First fixation durations at the post-target region showed no significant effect of grammaticality. See Figure 6 for means and standard errors.

2.2.8.2 *First pass reading times*.

First pass times showed a significant effect of grammaticality, F (2,114) = 3.27, MSE = 23675.25, $p < .05$, such that the grammaticality effect showed that only the noun-article condition (481 ms) had significantly shorter first pass reading times than the grammatical condition (520 ms), t (59) = 2.44, $p < .05$, but the noun-adjective condition (506 ms) was not significantly different from either of these. Refer to Figure 7 for means and standard errors.

2.2.8.3 *Regressions out*.

There was a significant effect of grammaticality, F (2,114) = 15.75, MSE = 0.53, $p < .001$, such that the post-target regions in the grammatical sentences had significantly fewer regressions (0.20) than in the ungrammatical noun-article sentences (0.30), t (59) = 3.27, $p < .01$, which in turn also had significantly fewer regressions than the ungrammatical noun-adjective sentences (0.39), t (59) = 5.24, $p < .001$. See Figure 8 for means and standard errors.

2.2.8.4 *Go past times*.

Go-past measures in the post-target region showed a significant effect of grammaticality, F (2, 114) = 13.30, MSE = 1484279.19, $p < .001$, such that the grammatical condition had the shorter go-past time (834 ms) than the ungrammatical noun-adjective condition (1156 ms), t (59) = 4.50, $p < .001$, and a marginal difference with the ungrammatical noun-article condition (970 ms), t (59) = 2.42, p = .02. There was also a significant difference between these two ungrammatical conditions, t (59) = 2.83, $p < .01$. See Figure 9 for means and standard errors.

2.2.8.5 *Total times*.

There was a significant effect of grammaticality, F (2,114) = 11.66, MSE = 1357840.29, $p < .001$, and post-hoc t-tests showed that the ungrammatical noun-article condition was viewed for significantly less total time (925 ms) than both the grammatical (1187 ms), t (59) = 4.27, $p < .001$, and ungrammatical noun-adjective (1184 ms), t (59) = 4.52, $p < .001$, conditions. There was no significant difference in total times between these last two. See Figure 10 for means and standard errors.

2.2.9 Target region summary of effects.

In summary, the target region shows different effects across the different measures. First fixation durations, as mentioned previously, are not the best measure for these analyses because of the nature of their measurement because it most likely only captures the fixations on the first two words of the target region. The analysis revealed that there was a significant effect of grammaticality, such that the ungrammatical noun-article condition was fixated for significantly longer than the grammatical and the noun-adjective conditions. This is most likely due to the participants being able to detect the ungrammaticality in the noun-article condition in the first two words of the target region. The first pass reading times show only a marginal effect of grammaticality, but a significant interaction with native language group: the native English speakers spent more time on both ungrammatical violations, but neither the native Arabic nor the native Mandarin speakers showed this difference. The native English speakers show a grammaticality effect, such that they do not process the two ungrammatical conditions differently during their first pass reading, but the native Arabic speakers showed a greater disruption for the noun-article condition with longer first pass reading times. In contrast, the native Mandarin speakers showed shorter first pass reading times than the grammatical and noun-adjective conditions. This could reflect differences in initial processing of the syntactic violations by the L2 speaker groups such that the native Arabic speakers show more disruption to the similar condition but the native Mandarin speakers show less disruption to the unique condition.

Regressions out measures showed this same pattern, the ungrammatical conditions had more regressions out of the target region than grammatical condition, but the noun-article condition had significantly more regressions out than the other two conditions. This demonstrated that this condition showed more disruption for processing across native language groups, indicating that the strength of the syntactic cue of article position may have influenced processing in a more substantial way than the similarity modulations. Go-past measures revealed all significant effects – the grammatical condition had the shortest go-past times compared to the ungrammatical conditions, and the noun-article target region had a significantly longer go-past time than the noun-adjective one, which could reflect the same type of cue strength effect seen from the regressions out. Lastly, all effects were again significant for total time spent on the region – the grammatical condition had the lowest total time spent on it, but an interesting interaction arose such that the native Mandarin and native English speakers displayed this grammatical v. ungrammatical distinction, but the native Arabic speakers only had significantly different total time for the noun-adjective condition and no difference between the noun-article and the grammatical condition. They did trend to show the same effect as the native English and native Mandarin speakers for a grammaticality effect between the noun-adjective and grammatical sentences, but the difference between the noun-article and grammatical condition did not reach significance. Total time measures index later syntactic processing and wrap up effects so this measure may show more processing similar to native patterns because L2 speakers have processed through to the wrap-up effects.

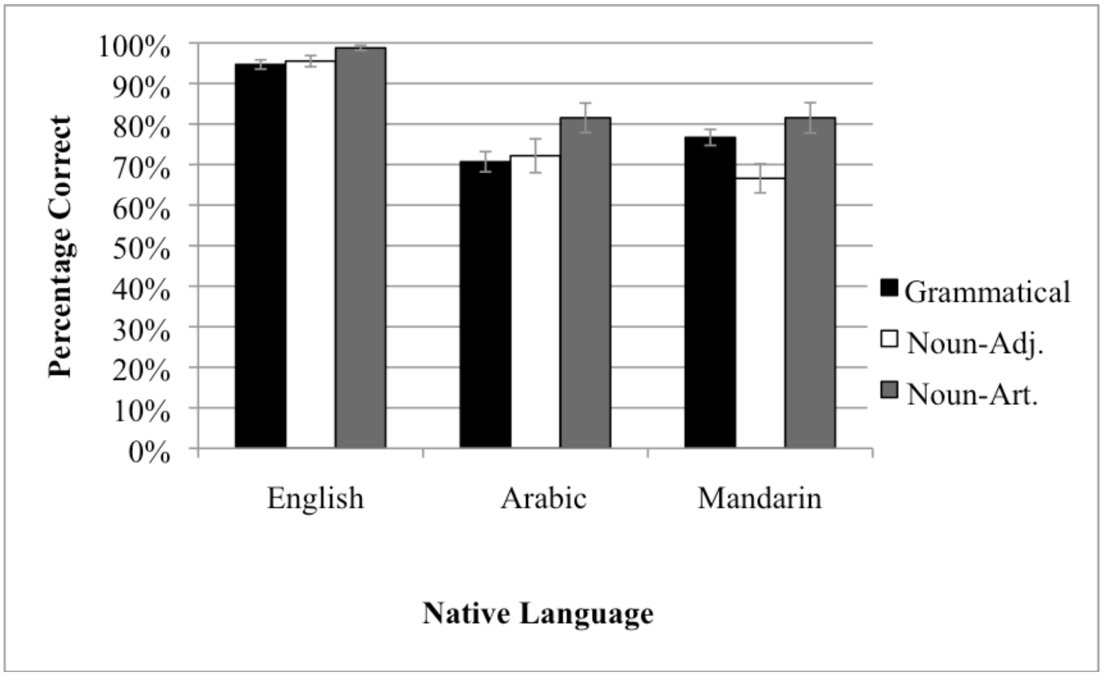
2.2.10 Post-target region summary of effects.

For the post-target region, there was a significant effect of language in the first fixation durations such that native Arabic speakers showed longer post-target first fixations than native English speakers. This was reflected in the first pass reading times as well – native Arabic and native Mandarin speakers showed longer first pass reading times than native English speakers. There was also a significant effect of grammaticality such that only the noun-article and the grammatical conditions were significantly different, showing spill-over effects from the strong syntactic cue in the critical region. Regressions out showed significantly fewer regressions for the grammatical post-target region, followed by the noun-article condition, and the noun-adjective region had the most regressions out. There was also a marginally significant interaction such that native Arabic speakers had more regressions out of the noun-adjective post-target region but the grammatical and noun-article conditions were not different from each other. The native Mandarin and native English speakers showed a grammaticality effect such that the grammatical post-target regions had fewer regressions out than both ungrammatical post-target regions. This reflects spill-over effects from the grammatical and ungrammatical critical regions: those post-target regions that were preceded by a grammatical critical region had less regressions out than the regions that were preceded by ungrammatical critical regions. Go-past measures for the post-target region showed a significant grammaticality effect, such that the grammatical condition had shorter go-past times than the ungrammatical conditions. There was also a significant interaction between grammaticality and native language group such that the native Mandarin speakers showed longer go-past times for the ungrammatical than grammatical conditions but no other group showed this significant difference; native Arabic speakers had longer go-past times for the noun-adjective sentences condition and both ungrammatical conditions were significantly different from each other for the native English speakers. Lastly, total time spent on the post-target region was analyzed and this showed a significant effect of grammaticality so that the noun-article condition was viewed for significantly less time than both the grammatical and the noun-adjective conditions. This last effect could be due to the stronger syntactic cue that precedes the ungrammatical noun-article condition. As soon as participants process the ungrammaticality completely in the critical region with the very strong syntactic cue, they know that it is ungrammatical and processing is aborted in later regions of the sentence. Although sentence-final effects were not analyzed, experimenters observed that participants would not read the full sentence once they had processed the violation, and go directly to answer the grammaticality judgment.

2.2.11 Grammaticality judgment task.

Grammaticality judgment accuracy showed a significant effect of grammaticality, F (2,114) = 9.44, MSE = 0.16, $p < .001$, such that the ungrammatical noun-article condition was judged marginally more accurately (86% accuracy) than the grammatical condition (81%), t (59) = 2.23, $p = .03$, but was judged significantly more accurately than the noun-adjective conditions (76%), t (59) = 5.21, $p < .001$, (see Figure 11). There was also a significant effect of native language group, F (2,57) = 28.21, MSE = 1.06, $p < .001$, and corrected post-hoc multiple comparisons showed that the native English participants had the best accuracy (96%) on the grammaticality judgment task compared to either the native Arabic (75%) ($p < .001$) or native Mandarin (72%) speakers ($p < .001$). The L2 speakers showed no differences between groups. There was also a significant interaction between grammaticality and native language group, F (4,114) = 2.70, MSE = 0.05, $p < .05$, and post-hoc t-tests demonstrated that the native English speakers had marginally better performance on the noun-article condition than the grammatical condition, t (19) = 2.47, $p = .02$. Native Arabic speakers had better accuracy on the noun-article condition as well compared to the noun-adjective condition, t (19) = 4.61, $p < .001$, and to the grammatical condition, t (19) = 2.96, $p < .01$. Lastly, native Mandarin speakers showed significantly worse performance for the noun-adjective condition than the noun-article condition, t (19) = 3.50, $p < .01$, and marginally worse than the grammatical condition, t (19) = 2.40, p = .03.

Figure 11. Grammaticality judgment task accuracy



2.2.12 Correlation analyses.

We were interested in the role of proficiency in the manner in which participants read sentences of various types because increased L2 proficiency should correspond to more native-like processing. Self-rated proficiency correlates highly with objective measures of proficiency (Blanche, 1988) and so these measures were used for the correlation analyses that follow. We performed correlational analyses based on the participants' self-rated proficiency in English across four measures (reading, writing, speaking and comprehension), four eye-tracking measures in the target region (first pass reading times, regressions out, go-past time, and total time), and their grammaticality judgment performance.

For the native English speakers, there was a significant correlation between self-rated proficiency in English reading and their performance in correctly judging the ungrammatical noun-article sentences (see Table 4), such that the higher they rated themselves on English reading skill, the better they were at judging grammaticality in the noun-article condition. All three total reading time measures had significant negative correlations with their self-rated English writing skills (see Table 5) such that the higher they rated themselves on English writing skills, the less total time they spent on the target region. No other significant correlations were found. See Tables 4 and 5 for a complete summary of the results by language group.

For the native Arabic speakers, self-rated English reading skill significantly correlated with their performance on judging the ungrammatical noun-article sentences (see Table 4) so the higher they rated themselves on their English reading skill, the better they performed when judging grammaticality for these sentences. No other significant correlations were found.

For the native Mandarin speakers, self-rated English comprehension and reading skills correlated significantly with judging both types of ungrammatical sentences (see Table 4) and these same skills significantly negatively correlated with all go-past times for all three types of sentences such that the higher they rated themselves on their comprehension and reading, the better they performed during the grammaticality judgment task. Self-rated English reading skills correlated significantly with first pass reading times, but not with any other skill. Self-rated English conversation and writing skills also significantly negatively correlated with go-past times for the noun-adjective sentences and English conversation correlated significantly with go-past times for the noun-article sentences so the higher they rated themselves on these, the lower their duration on these regions for go-past times (see Table 5). No other significant correlations were found.

Table 4. *Correlations between participants' English skills and grammaticality judgment task.*

	English Reading	English Writing	English Conversation	English Comprehension
ENGLISH				
Accuracy - Gramm.	-.05	-.10	.07	.25
Accuracy - Noun-Adj.	-.22	-.28	-.27	-.19
Accuracy - Noun-Art.	.46*	.27	.04	.30
MANDARIN				
Accuracy - Gramm.	.16	.10	-.13	-.23
Accuracy - Noun-Adj.	.43*	.33	.26	.42*
Accuracy - Noun-Art.	.36*	.33	.34	.50**
ARABIC				
Accuracy - Gramm.	-.10	.09	.20	.03
Accuracy - Noun-Adj.	.34	.23	-.07	.06
Accuracy - Noun-Art.	.37*	.22	-.05	.03

** Correlation is significant at the 0.01 level (2-tailed).

* Correlation is significant at the 0.05 level (2-tailed).

Table 5. *Correlations between participants' English skills and reading time measures.*

	English Reading	English Writing	English Conversation	English Comprehension
ENGLISH				
First Pass Gramm	-.28	-.29	-.04	-.09
First Pass Noun-Adj.	-.23	-.28	.02	-.13
First Pass Noun-Art	-.25	-.29	.01	.03
Regr. Out Gramm	-.18	-.37	-.09	.07
Regr. Out Noun-Adj.	-.06	-.09	.05	.28
Regr. Out Noun-Art	-.23	-.34	.07	.15
Go-past Gramm.	-.25	-.22	.02	.04
Go-past Noun-Adj.	-.29	-.31	.00	-.03
Go-past Noun-Art.	-.09	-.08	.08	.10
Total time Gramm.	-.33	-.46*	-.09	.01
Total time Noun-Adj.	-.31	-.45*	-.08	-.02
Total time Noun-Art.	-.26	-.38*	.00	.00
MANDARIN				
First Pass Gramm	-.37*	-.23	-.15	-.19
First Pass Noun-Adj.	-.39*	-.22	-.24	-.22
First Pass Noun-Art	-.40*	-.27	-.29	-.11
Regr. Out Gramm	-.21	-.15	-.13	-.17
Regr. Out Noun-Adj.	-.13	-.29	-.20	-.14
Regr. Out Noun-Art	-.11	-.15	-.05	-.18
Go-past Gramm.	-.47**	-.33	-.29	-.40*
Go-past Noun-Adj.	-.44*	-.39*	-.39*	-.46**
Go-past Noun-Art.	-.40*	-.32	-.36*	-.49**
Total time Gramm.	-.31	-.20	-.07	-.10
Total time Noun-Adj.	-.12	-.16	-.03	-.11
Total time Noun-Art.	-.27	-.31	-.14	-.16
ARABIC				
First Pass Gramm	-.30	-.30	-.31	-.33
First Pass Noun-Adj.	-.24	-.13	-.10	-.17
First Pass Noun-Art	-.03	-.03	-.03	.18
Regr. Out Gramm	.06	.05	.14	.16
Regr. Out Noun-Adj.	.15	.27	-.03	.02
Regr. Out Noun-Art	.11	.01	.03	.17
Go-past Gramm.	.05	-.06	-.17	-.29
Go-past Noun-Adj.	-.14	-.01	-.16	-.27
Go-past Noun-Art.	-.10	-.08	-.07	-.34
Total time Gramm.	-.18	-.03	-.06	-.11
Total time Noun-Adj.	-.18	.01	-.06	-.17
Total time Noun-Art.	-.10	.00	-.05	-.18

** Correlation is significant at the 0.01 level (2-tailed).

* Correlation is significant at the 0.05 level (2-tailed).

In summary, participants had significantly better performance on the ungrammatical noun-article sentences than the XX sentences in the grammaticality judgment task, and only native Mandarin speakers showed similar performance between the grammatical and the noun-article sentences. The correlations demonstrate that self-rated English reading skills across language groups correlates significantly with accuracy judgments on the noun-article sentences, and that native Mandarin speakers also have significant correlations between English comprehension skills and their performance on both types of ungrammatical sentences such that the higher they rated themselves on their English skills, the shorter their go-time durations and the better their performance on judging grammaticality.

3.0 DISCUSSION

Overall, the online measures demonstrated different levels of processing as L2 speakers read the sentences with two types of syntactic violations. We will not concentrate on the first fixation durations due to the parafoveal confound (elaborated more in detail below), but the other four measures show different stages of processing that are contingent upon each other and may reflect how L2 learners resolve syntactic violations online.

First fixation durations demonstrated that all speaker groups had longer first fixations for the noun-article condition than either of the other two, and this may be explained by the fact that the first word marks a violation in the noun-article condition as opposed to the other two.

First pass reading times for the non-native speakers showed no differences between the conditions, and the native English speakers showed a grammaticality effect such that both ungrammatical sentences had longer first pass reading times than the grammatical. Native Arabic speakers had longer reading times for sentences in the ungrammatical noun-article condition than the other two conditions but native Mandarin speakers showed shorter reading times for sentences in the the noun-article condition. As the first full pass through the critical region, this measure may reflect the initial processing that takes place as the participants encounter the violation for the first time. Native English speakers the three types of sentences with no differences, but native Arabic speakers showed longer times for noun-article condition, which was our similar condition. Therefore, they showed increased sensitivity to the similar condition, but native Mandarin speakers showed the opposite effect, in that they had less disruption to the noun-article condition. This could be due to the fact that Mandarin does not have a true definite article so they do not pay as much attention to it across their first processing sweep.

Similar to the first fixation results, regressions out showed more regressions for the noun-article sentence than either the noun-adjective or grammatical conditions across all three language groups. More regressions out of a region are indicative of more disruption in processing that critical region. As the participants are reading across the sentence, the noun-article condition disrupts reading more than the other conditions and they regress more out of the region to presumably reread the sentence over again and try to comprehend it more. Regressions out builds on the first pass reading times, so it seems that after the first sweep through the critical region, the noun-article sentence is the most difficult to process. This is possibly due to the fact that the syntactic cue of the article violation is much stronger than that of the other sentences and is the cause of more difficulty and regressions out. Looking at later measures will allow us to better qualify these conclusions.

Native English speakers showed sensitivity at both later measures (go-past times and total time) with a bigger difference between the noun-article violations and the grammatical condition than the noun-adjective region. This same pattern was shown by both non-native speaker groups but at different measures: the native Arabic speakers showed the same effect as the native English speakers in go-past time, whereas the native Mandarin speakers showed a grammaticality effect at the total time measure such that both types of violations had longer total times than the grammatical condition.

The results of the non-native speaker groups in these measures demonstrated patterns similar to those of the native speakers for syntactic violations in later reading time measures, which does not conform to the predictions put forth by the SSH. The model put forth by Clahsen and Felser (2006a) makes a sharp distinction between morphosyntactic and pure syntactic resolution and the effects that an L1 may have on these. Native-like processing cannot be achieved for complex syntactic structures because the available L2 grammar does not provide sufficient syntactic information to resolve these structures. However, the results of this study demonstrate that native-like processing can be seen for non-native speakers at later times in processing. The native Arabic speakers showed the same magnitude of disruption in the go-past measures as the native English speakers and the native Mandarin speakers demonstrated similar results in their total time measures as the native English speakers. This may due to the fact that the syntactic structures presented in this study were not as complex as those addressed by the SSH, which dealt primarily with non-local syntactic dependencies. Local syntactic and morphosyntactic dependencies may well reach native-like processing, so the results of this study provide some support for this statement. As Clahsen and Felser (2006b) point out, what remains to be seen for the SSH is whether late learners of an L2 can process purely syntactic structures in a native-like way. This study addressed this gap in research by showing that non-native speakers can, in fact, exhibit processing patterns similar to those of native speakers when presented with syntactic violations in their L2. When confronted with structures that were similar, different, or unique in their construction to the participants' L1, non-native speakers did not exactly conform to the hypothesized modulations that the UCM predicted, but they did show variation between the conditions. This was, most likely, due in large part to the effects of the highly reliable syntactic cue of the noun-article condition. This construction was strong enough to elicit patterns across all three language groups that were similar, even though the construction in their first language may have prompted other possible interpretations (such as we predicted might have been the case for the native Mandarin speakers).

The UCM predicts that structures that are similar across languages, in this case the position of the article for the native Arabic and native English speakers, show positive transfer from L1 to L2, but those that are different (the noun-adjective sentences) cause competition between the competing syntactic cues from each language. Therefore, the native Arabic speakers should show less sensitivity in the noun-adjective condition and more sensitivity in the noun-article condition. In this way, the noun-article condition that was unique for the native Mandarin speakers would not have competition or transfer, but the similar condition (noun-adjective) would show transfer from the L1 to the L2, predicting that these speakers may be more sensitive to the noun-adjective violations. The go-past times for the native Arabic speakers showed a grammaticality effect and the magnitude between the noun-article and grammatical sentence was significantly more than the difference between the noun-adjective and the grammatical sentences, demonstrating that they showed more disruption to the noun-article violation than the noun-adjective violation, conforming to UCM predictions that this structure would show more sensitivity. This result is the same as the native English speakers' results – native Arabic speakers were more sensitive to the similar violation than to the different violation. Contrastingly, the native Mandarin speakers were similarly sensitive to both types of violations in the total time measure with no difference in sensitivity for the different similarity conditions.

A compelling argument for the results found here comes in the form of looking at the strength of the syntactic cues that were presented. Because the noun-article sentence has a violation that provides a much more reliable and strong cue than the noun-adjective sentence, it can account for the strong results across the three language groups and the similar processing results that were seen between the native English speakers and the L2 speakers. The very strong cue provides information that even non-native speakers can use to readily process ungrammatical sentences in much the same way as native speakers do. This may provide a new dimension for models such as the UCM because there may be syntactic cues that are so strong that they override effects of similarity that transfer models, such as the UCM, predict. Controlling for cue strength and reliability, therefore, should be attended to in future work because it can readily influence how L2 speakers process syntax, such that they can show patterns like native speakers do during processing.

Another interpretation is that these results provide evidence against the SSH's notion of no transfer because the native Arabic speakers showed native-like processing in their go-past times such that their results mirrored those of the native English speakers, but they also showed more disruption to the noun-article violation, the similar condition across the two languages. Interestingly, in their total time measures, native Arabic speakers had marginally longer total time for the noun-adjective condition, demonstrating that processing the noun-adjective condition was more difficult than processing the noun-article condition. Although there was no significant difference in the magnitude of this effect, it is interesting to note the change from go-past time to total time in native Arabic speakers. They showed native-like processing during their go-past time, mirroring the native English speakers' results, but this effect was no longer present in total time. It may be that during go-past times, native Arabic speakers are able to process the syntactic violations in a native-like way, but once they go back to the target region after reading through the entire sentence, competition arises for the total time measures. It would be worthwhile to delve further into this contrast in reading time measures for L2 learners because it may be that different types of processing occur at different time points during the processing of an L2, as these results seem to suggest. However, the similarity modulation would be better supported if the native English speakers had not shown differences between the two ungrammatical conditions.

The clear differences between the different eye-tracking measures also provide valuable insight into how L2 speakers process and resolve syntactic violations in their L2. As L2 learners read across sentences in their L2, they may have immediate transfer from their L1, as could be interpreted from first pass reading times, but as they read more in-depth, they can attend to strong syntactic cues like native speakers. This was seen here during regressions out and go-past times, and even in their total time spent on the critical region, where the patterns of eye movements were very similar to those of native English speakers. Therefore, with enough of syntactic cue strength, L2 processing may behave just like native-like processing and this can be reflected in the different, but contingent, eye-tracking measures.

Finally, another facet of this thesis is that the similar results between the native and non-native speakers may be due to a higher level of L2 proficiency across the non-native speakers such that they showed a tendency to process these syntactic structures in a more native-like way.

The majority of previous studies have tested non-native participants with moderate to high levels or proficiency in their respective L2s. As those studies have also demonstrated, results vary with proficiency and the type of construction tested. For example, highly proficient L2 Greek speakers did not show native-like resolution preferences (Papadopoulou & Clahsen, 2003), but advanced L2 speakers of Dutch showed transfer from various L1s for gender markings (Sabourin et al., 2006). Moreover, transfer from the L2 back to the L1 has also been demonstrated with proficient L2 speakers of English and Spanish (Dussias, 2003), providing compelling evidence for models of transfer. Proficiency plays an important role in how L2 learners process different aspects of L2 and in this case, it could be that the moderate level of proficiency of the non-native speaker groups led them to process the syntactic violations in a native-like manner in later syntactically-modulated measures. This could especially be the case for the native Mandarin speakers who demonstrate native-like processing across both later reading time measures. As mentioned above, a portion of the participants had already been accepted into a university environment and so their objective English skills may have been superior to those of the native Arabic speakers who were all students in a specialized program to teach English to foreign students. The correlations we ran with their self-rated skills and their performance on the eye-tracking measures demonstrated that the native Mandarin speakers showed the most correlation between the two, with various skills correlating with various measures. However, neither the native Arabic nor native English speakers produced conclusive results. Future objective measures of proficiency may be of greater use here in parsing out the influence of different skills more specifically than self-ratings. However, the high degree of correlations between the native Mandarin speakers and their eye-tracking measures may be able to show why their performance was more similar to that of native speakers of English than native speakers of Arabic.

The results of this study provide important considerations for models of L2 processing such as the SSH and the UCM. The SSH posits a strict division between the processing of an L1 and an L2, such that non-native speakers do not have the necessary syntactic representations to appropriately parse syntactic violations in L2. If the two systems develop independently of one another, then native-like processing, such as seen here, cannot be achieved for purely syntactic constructions, contrary to their suppositions. The results of this study, as well as others, have shown that this is not the case. The non-native speakers demonstrated that native-like processing could arise at later reading time measures such that they mirror native English participants' results. These results can be better explained in the context of a model that allows competition between languages to influence processing through negative or positive transfer, as well as considerations for factors such as syntactic cue strength, much like the UCM posits. The native Arabic participants showed results that mostly conformed to the UCM predictions, in that they were more sensitive to the similar (noun-article) condition than the different (noun-adjective) condition. These participants and the native Mandarin participants showed native-like processing that a model like the SSH should try to account for. It may be necessary to revise these models of L2 processing and transfer to include factors such as syntactic cue strength or even more levels of proficiency. Once an L2 speaker has been given a strong and reliable cue to help them process the L2, then it may be that they show native-like processing to the degree that the similarities and differences that the UCM predicts can be overshadowed such that there are no differences in the similarity conditions. On the other hand, allowing for transfer to occur during the learning of an L2 and conceding that native-like processing in syntactic structures can take place in L2 learners may be the next step for models such as the SSH.

BIBLIOGRAPHY

Bauman, J. (2010). About the General Service List. Retrieved September 11, 2010, from http://jbauman.com/aboutgsl.html#1953

Blanche, P. (1988). Self-assessment of foreign language skills: Implications for teachers and researchers. *RELC Journal*, *19*, 75-93.

Chen, L., Shu, H., Liu, Y., Zhao, J., Li, P. (2007). ERP signatures of subject-verb agreement in L2 learning. *Bilingualism: Language and Cognition, 10*, 161-174.

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

Clahsen, H. & Felser, C. (2006b). How native-like is non-native language processing? *Trends in Cognitive Science*, *10*, 564-570.

Clifton, C., Jr., Staub, A., & Rayner, K. (2007). Eye movements in reading words and sentences. In R. van Gompel (Ed.), *Eye movements: A window on mind and brain* (pp. 341-372). Amsterdam: Elsevier.

Dussias, P. (2003). Syntactic ambiguity resolution in L2 learners: Some effects of bilinguality on L1 and L2 processing strategies. *Studies in Second Language Acquisition*, *25*, 529-557.

Dussias, P. (2010). Uses of eye-tracking data in second language sentence processing research. *Annual Review of Applied Linguistics*, *30*, 149-166.

Dussias, P. E., & Sagarra, N. (2007). The effect of exposure on syntactic parsing in Spanish English bilinguals. *Bilingualism: Language and Cognition*, *10*, 101–116.

Frenck-Mestre, C., & Pynte, J. (1997). Syntactic ambiguity resolution while reading in second and native languages. *Quarterly Journal of Experimental Psychology, 50*, 119–148.

Ju, M., & Luce, P. A. (2004). Falling on sensitive ears. *Psychological Science*, *15*, 314-318.

Just, M. A., Carpenter, P. A., & Woolley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General*, *111*, 228-238.

MacWhinney, B. (2005). A unified model of language acquisition. In J.F. Kroll & A.M.B DeGroot (Eds.), *Handbook of Bilingualism: Psycholinguistic Approaches* (pp. 49-67): Oxford University Press.

MacWhinney, B. & Pleh, C. (1988) The processing of restrictive relative clauses in Hungarian. *Cognition*, *29*, 95-141.

Marian, V., & Spivey, M. (2003). Bilingual and monolingual processing of competing lexical items. *Applied Psycholinguistics*, *24*, 173–193

Osterhout, L. & Nicol, J. (1999). On the distinctiveness, independence, and time course of the brain response to syntactic and semantic anomalies. *Language and Cognitive Processes, 14*, 283-317.

Papadopoulou, D. & Clahsen, H. (2003). Parsing strategies in L1 and L2 processing: A study of relative clause attachment in Greek. *Studies in Second Language Acquisition*, *25*, 501-528.

Picton, T.W., Bentin, S., Berg, P., Donchin, E., Hillyard, S.A., Johnson, Jr., R., Miller, G. A., Ritter, W., Ruchkin, D.S., Rugg, M.D., & Taylor, M.J. (2000). Guidelines for using human event-related potentials to study cognition: Recording standards and publication criteria. *Psychophysiology*, *37*,127–152.

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology, 62*, 1457-1506.

Rogerson, H. D., Davis, B., Hershelman, S. T., & Jasnow, C. (1992). *Words for students of English: A vocabulary series for ESL, Vol. 1*. Ann Arbor, MI: University of Michigan Press.

Rogerson, H. D., Esarey, G., Hershelman, S. T., Jasnow, C., Moltz, C., Schmandt, L. M., & Smith, D. A. (1992). *Words for students of English: A vocabulary series for ESL, Vol. 4*. Ann Arbor, MI: University of Michigan Press.

Rogerson, H. D., Hershelman, S. T., Jasnow, C., & Moltz, C. (1992). *Words for students of English: A vocabulary series for ESL, Vol. 3*. Ann Arbor, MI: University of Michigan Press.

Sabourin, L. & Stowe, L. (2008). Second-language processing: when are first and second languages processed similarly? *Second Language Research, 25, 3*, 397-430.

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

Spivey, M., & Marian, V. (1999). Cross talk between native and second languages: Partial activation of an irrelevant lexicon. *Psychological Science, 10*, 281–284.

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

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

Tokowicz, N. & Warren, T. (2010). Beginning adult L2 learners' sensitivity to morphosyntactic violations: A self-paced reading study. *European Journal of Cognitive Psychology*.

Tolentino, L. & Tokowicz, N. (2011). Across Languages, Space and Time: The role of cross-language similarity in L2 (morpho)syntactic processing as revealed by fMRI and ERP. *Studies in Second Language Acquisition, 33*, 91-125.

Warren, T. (2011). The influence of plausibility and anomaly on eye movements in reading. In S. P. Liversedge, I. Gilchrist, & S. Everling (Eds.), The Oxford Handbook on Eye Movements (pp. 911-923). USA, Oxford University Press.

Warren, T., McConnell, K., & Rayner, K. (2008). Effects of context on eye movements when reading about possible and impossible events. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 34*, 1001-1010.

West, M. (1953). *A general service list of English words*. London, UK: Longman.

Witzel, J., Witzel, N., & Nicol, J. (2009, May). The reading of structurally ambiguous sentences by English language learners. Paper presented at the Workshop on Second Language Processing and Parsing: State of the Science, Lubbock, TX.