RESOLUTION OF GENDER CUE CONFLICT BY ARABIC-SPEAKING LEARNERS OF ENGLISH: AN EYE-TRACKING STUDY

by

ELENA DECOOK

DILIN LIU, COMMITTEE CHAIR
ROBERT POOLE
LAURA MORETT

A THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Arts
in TESOL/Applied Linguistics
in the Graduate School of
The University of Alabama

TUSCALOOSA, ALABAMA

2020

ABSTRACT

While every English learner experiences the learning process differently, some phenomena are more common for learners from specific L1 groups. In the case of this study, speakers of Arabic as an L1 often find that reading in English presents a level of difficulty beyond that of speaking, listening, or writing. Many theories have been suggested to explain this occurrence, especially theories that hinge on the differences between the writing systems and phonology of the two languages. Research on the differences between Arabic L1 and English L1 readers on the syntactic level is less common, but information about higher-order processes is no less essential to understanding how Arabic L1 English learners approach reading.

This study was conducted in order to better understand how this learner population processes written text in English. This procedure examines how Arabic L1 and English L1 readers react to gender cue mismatching in written materials by using eye tracking technology. Specifically, the study consists of an experiment that follows the guidelines of Kreiner, Sturt & Garrod (2008) to assess how Arabic L1 learners of English are affected by gender mismatches between reflexive pronouns and the nouns they reference. Sentences were presented in 2x2 conditions: the subject of each was either a definitionally-gendered noun, like "queen," or a stereotypically-gendered one, like "nurse," and each sentence contained a reflexive pronoun that either matched or mismatched the gender of the noun. An infrared eye-tracking camera was used to determine the location of participants' gazes over time while reading. The resulting data was analyzed to determine the effects of both conditions on Arabic L1 English learners in contrast to native English speakers.

Results indicate that Arabic L1 English learners are affected by gender cue conflict, but the processing costs incurred by resolving that conflict are incurred later in the reading process in comparison to L1 English speakers. This is consistent with eye-tracking results derived from other non-native English readers and presents new evidence of processing differences on the syntactic level between Arabic L1 and English L1 readers. Implications for pedagogy and future research are discussed.

LIST OF ABBREVIATIONS AND SYMBOLS

ANOVA	analysis of variance
COCA	Corpus of Contemporary American English
EEG	electroencephalogram
ESL	English as a second language
fNIRS	functional near-infrared spectroscopy
IELTS	International English Language Testing System
L1	first language
L2	second language
NL	native language
NS	native speaker
NNS	non-native speaker
TOEFL	Test of English as a Foreign Language
TOEFL iBT	Test of English as a Foreign Language internet-based test

ACKNOWLEDGEMENTS

Firstly, I would like to thank the members of my thesis committee, Dr. Dilin Liu, Dr. Laura Morett, and Dr. Robert Poole. Without their incredibly generous help, none of this research would have been possible. Their expertise, guidance, and feedback made up the backbone of this project. In particular, I would like to thank the chair of my committee, Dr. Dilin Liu, for providing endlessly patient advice and structure over the past year.

Secondly, I would also like to thank the NERD lab at the University of Alabama, for allowing the use of their eye-tracking equipment, and in particular Sarah Hughes Berheim, Jake Feiler, and Cailee Nelson for kindly answering my dozens of data-gathering questions.

Thirdly, I would like to thank all of the teachers who allowed me to recruit from their composition classes. I would also like to thank the students who participated in this research project for their time, energy, and open-mindedness.

Finally, I would like to thank all of the many others who supported me as I completed this project: my friends and family, my cat, and the bus drivers on the Silver 2 line. This thesis is much less a product of my own work and ideas than it is a product of the help I received this year, and I cannot express my thanks enough.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

Defining the Issue

Understanding language cognition is central to understanding language acquisition (Winke, Godfroid & Gass, 2013). In many ways, successful teaching and learning of a second language depends on knowledge about the implicit, near-immediate processes that underlie comprehension of meaning. However, due to those factors of speed and control, traditional "offline" linguistic methods, such as think-aloud protocol, cannot always offer accurate processing data, as they rely on tasks that assess processing after the fact (Conklin & Pellicer-Sánchez, 2016). For applied linguists and language teachers, who aim to facilitate language development by grounding their pedagogy in theory about how that development occurs, more direct, real-time "online" methods of measurement and the greater accuracy they offer about processing are particularly appealing.

One such methodological option is eye-tracking. Use of this technique is based on a construct now called the "eye-mind assumption," defined by Just & Carpenter (1980). That is, eye-tracking researchers believe that covert attention—i.e., the location of a viewer's gaze—is directly reflective of and connected to overt attention, or the information which the viewer is processing at the same moment. Because it can record changes in eye-gaze down to the millisecond and the 1/100th of a degree, and because it does not require the reader to perform a secondary task while or after reading, eye-tracking is increasingly common in second language acquisition (SLA) studies, contributing essential information about the myriad of ways in which processing occurs before learners can provide a response to a stimulus (Conklin & Pellicer-Sánchez, 2016, Roberts & Siyanova-Chanturia, 2013, Winke, Godfroid & Gass, 2013).

In particular, eye-tracking has been used to gauge, among other things, the ways that bilingual individuals navigate written text (Winke, Godfroid & Gass, 2013) and process multiword units (Siyanova-Chanturia, 2013). These reading studies, which are equally popular in first language (L1) research, have investigated how readers parse individual words, bind referents together, apply conscious reading strategies, among other foci. The topics of the present study—that is, how learners incorporate lexical information and a reflexive structure in order to comprehend a sentence—have been previously addressed in both L1 literature (Kreiner, Sturt & Garrod, 2009) and SLA literature (e.g. Felser, Sato & Bertenshaw 2009, Felser & Cunnings 2012). This line of research has provided many insights into how readers retrieve lexical knowledge, what effects ambiguity can have on reading speed and comprehension, and how non-native English speakers (NNS) process English sentences differently from their native English-speaking (NS) counterparts. This information has many potential ramifications for the learning and teaching of reading skills to students of English as a second language.

This research, however, has been rarely conducted with NNSs from Arabic-speaking backgrounds (Abu-Rabia, 1997)—a clear oversight, given the well-documented difficulties that Arabic L1 students have with reading in English (Alhazmi, Milton & Johnson, 2019; Alsadooh & Heift, 2015; Clahsen, Balkhair, Schutter & Cummings, 2013; Jordan et al., 2014; Tuninetti, Warren, & Tokowicz, 2015). Research has suggested many reasons for the laboriousness of this reading process for Arabic ESL/EFL learners, including the stark visual differences between the writing forms of the two languages, the distinctions in their grammar and orthography, and a possible transfer of reading strategies between languages, but conclusive evidence for one theory over another is sparse. This is not a niche issue, especially not for SLA professionals in the United States. During the 2018/2019 academic year, at least 5.3% of all international students in the United States spoke Arabic as a first language, and Saudi Arabia was the fourth most-popular

country of origin for international students overall, contributing more than 37,000 students to U.S. campuses (The Institute of International Education). The state of Alabama hosted a still-higher proportion of Arabic speakers in 2019: led only by China and India, Saudi Arabia and Kuwait made up 9.3% of Alabama's international student population (The Institute of International Education). The University of Alabama itself is no exception: of all international students actively enrolled at UA for the 2019/2020 academic year, 39.12% (568 of 1,452) were from the Middle East & North Africa: 427 from Kuwait, 81 from Saudi Arabia, and 31 from Iran, among other Arabic-speaking countries like Oman and Yemen (The University of Alabama). As part of their university studies, this group will be expected to read and absorb a great deal of written material, but many questions still remain about the unique ways in which they read. Thus, especially at the University of Alabama, there is both the distinct need and the valuable opportunity for further study of how native Arabic speakers process language while reading in English.

Chapter Overview

The report of the present study is organized into six chapters. The first chapter introduces the basis and goals of the study. The second chapter includes reviews of relevant literature. First, this chapter summarizes patterns in research on Arabic L1 English learners, particularly as they relate to reading processes. Second, this chapter provides an overview of eye-tracking findings relevant to reading and cognition. Finally, it describes basic principles of syntactic binding which form context for the cognitive procedures used in this study. The third chapter outlines the methodology used in this study, beginning with a description of the experiment's participants, then explaining the production process of the experimental materials, and finally describing the procedure of the present study. The fourth chapter reports quantitative data derived from the recordings and compares findings across demographic groups. Chapter 5 presents a more

detailed analysis of said data, synthesizing several noticeable patterns to form a theory about Arabic L1 English cognition during reading and the unique challenges Arabic ESL students experience in reading. Finally, Chapter 6 relates the implications of the present study, including its limitations, suggestions for future study, and possibilities for the effect of these results on the ESL or EFL classroom.

CHAPTER 2: LITERATURE

This chapter presents a base of literature in several areas: firstly, the challenges that Arabic-speaking learners of English face when learning to read in their L2, including vowel blindness and other considerations; secondly, the previous and potential uses of eye-tracking for reading research both outside and inside of second-language acquisition studies; and thirdly, the principles and previous eye-tracking research concerning anaphors, reflexives, and coreferent binding. Studies are then synthesized to outline the purposes of the study described in this thesis, and research questions which guide the study are posed.

Reading Challenges Faced by Arabic-Speaking Learners of English

Interference

While theories of strong language interference are no longer as definitive in language teaching as they once were, it is still clear that a learner's first language has significant effects on their language-learning process (Roberts, Gullberg & Indefrey, 2008; Sagarra & Ellis, 2013; Tuninetti, Warren, & Tokowicz, 2015; Van Assche, Duyck, & Brysbaert, 2013). A learner's first language, and the differences between it and their target language, can affect the process of L2 learning enormously. Research has shown such L1 influence or interference is especially strong for Arabic speakers in learning English, especially in learning to read English (Thompson-Panos & Thomas Ružić, 1983), beyond the level of difficulty that language-learners from other groups experience. Most theories about this phenomenon have to do with differences between Arabic and English in more mechanical aspects of reading skills, like word-recognition (Fender 2003) and phonological parsing. There are a few obvious differences between the two written

languages on the surface: the Arabic and Latin alphabets use entirely different symbols and connection schemes, and Arabic is written from right to left, the opposite direction of English. While there is no universally-accepted explanation for the difficulties of Arabic L1 English learners in reading, some have been investigated more thoroughly than others.

One established theory in this area is called "vowel blindness." Because Arabic writing often only includes vowels as diacritics, and because the same information inferred from vowels can also be inferred by context clues, such as the aspect of nearby nouns or the conjugation of nearby verbs, it is believed that native speakers of Arabic learn to ignore vowels when parsing a word (Alsulaimani, 1990; Ryan & Meara, 1991). Because so much information is encoded in English vowels, this predictably causes processing difficulties for Arabic L1 English learners.

Vowel blindness

Two recent studies use eye-tracking to probe the effects, presentation, and potential interventions of vowel blindness. Alhazmi, Milton & Johnston (2019) investigate the word-recognition processes used by Arabic L1 English learners and, in particular, whether it is possible that the resulting reading difficulties are the product of cross-linguistic transfer—that is, whether Arabic L1 English learners borrow reading strategies from their L1 which interrupt reading in their L2. While the authors concede that vowel blindness is a well-established phenomenon with Arabic learners of English, this study found that when Arabic speakers read decontextualized English words, they did not linger on consonants more than on vowels. These results are limited, however, by the fact that they did not embed the words in sentences. If, as Alsulaimani (1990) suggests, native speakers of Arabic tend to focus on context clues to gather information otherwise encoded in vowels, then it is perfectly logical for an Arabic L1 English learner to focus on all letters in a decontextualized word. While vowel blindness is perhaps not

universal to all reading situations for Arabic L1 English learners, it would logically appear in readings of a sentence or longer.

Another study by Alsadooh & Heift (2015) attempts to address interventions for readers who struggle because of vowel blindness. Their study involved phonetic awareness training for beginner ESL learners from Arabic L1 backgrounds: target words which required vowel decoding for comprehension were visually emphasized via highlighting, italicizing, and bolding. They used an eye-tracker to record gaze movements during the treatment phrase, and researchers found that this treatment did, in fact, reduce the occurrence of vowel blindness in students who underwent training. Based on the eye-tracking data, they theorize that the longer fixations caused by textual enhancement led to better comprehension of the target words. This research suggests that vowel blindness can be addressed in SLA pedagogy in order to improve reading abilities in Arabic-speaking English learners.

Other recorded patterns in Arabic L1 English reading

While research in this area was once called relatively sparse (Abu-Rabia, 1997), Arabic L1 learners' reading development has become an increasingly popular phenomenon in SLA eye-tracking research. While vowel blindness is a frequently-investigated theory, it is by no means the only topic approached by researchers in Arabic L1 English acquisition. Clahsen et. al (2013), for instance, tested the effects of masked priming on Arabic L1 learners' ability to read morphologically complex words. This priming occurred by presenting a word related to the target word to the reader for 60 ms, a length of time which is too short for readers to actually identify the word, but which has been found in native English speakers to provide enough implicit information to allow them to process the target word more quickly when it appears. For Arabic-speaking English learners, however, priming did not result in faster processing, not even when they were presented with the prime for a lengthier period of time. Clahsen et. al conclude

that the Arabic speakers, who did not benefit from longer priming times, likely did not begin grammatical parsing as immediately as native English speakers.

Other studies, however, point out the similarities in Arabic and English reading. The work of Jordan et. al (2014) examines the perceptual span—i.e., the "moving window" of text from which a reader can extract useful information during a reading fixation—of Arabic readers. The English "moving window" is asymmetrical, with a reader's awareness extending further to the right of the point of fixation, and usually extends .5 degrees to the left and 1.5-2.5 degrees to the right. Jordan et. al found that the perceptual span for Arabic readers was a similar size and possessed a similar degree of asymmetry: the only difference was the direction toward which the asymmetry occurred. While English reading moves from left to right, and therefore the perceptual window extends further into unread territory, which is on the right, Arabic reading moves from right to left, and thus the perceptual window extends further into the unread territory on the left. Other maxims of eye-tracking reading research in English hold true for the same type of investigation into Arabic. For instance, one of the most predictable aspects of a reader's eye movements is that a longer word will garner more and longer fixations. Paterson et. al (2015) find that, despite the orthographic differences in the two languages, this effect is identical in English and Arabic. While eye-tracking has recorded significant differences in processing patterns for Arabic L1 and English L1 readers, it has also shown that the processes of reading Arabic and English are not too disparate to be compared.

Utility of Eye-Tracking in Research on Reading

General reading studies in eye-tracking

Eye-tracking for the purpose of linguistics research has been in practice since as early as 1906 (Dearborn). The theoretical basis for this, which is now commonly called the "eye-mind assumption," supposes that a reader's gaze is usually fixed on the word or word segment which it

is attempting to process (Just & Carpenter, 1980). Unlike a listener, who cannot control the rate of input, a reader can pause over stimuli that they require more time to understand. Words might attract longer fixations if they require sight-reading, parsing and retrieving the meaning of a rare word, choosing among several possible meanings, linking a word to a referent, or resolving ambiguities about its place in the discourse. By tracking the movements of a reader's eye around a sentence, researchers can begin to understand the way that the sentence is being processed.

There are many advantages to such a method. It has unique applications even among other methods that directly record brain activity. While it is possible to measure actual brain activity with, for instance, fNIRS or EEG machines, eye-tracking has the benefit of occurring in a physically naturalistic setting (Chun, 2003). Even if their head movements are controlled by bite bars or chin rests, readers are able to sit normally and move their arms freely, they can leave the recording setup without assistance from another person, and unless the eye-tracker is head-mounted, they do not wear any garments related to recording. Certain eye-trackers are also designed to be portable, so that researchers can, for instance, go to an elementary school to record eye-tracking data from children (Dussias 2010). This leads to data that is more representative of readers' regular behavior.

The variety of data available also contributes to the flexibility of eye-tracking applications. Eye trackers report three types of phenomena: movements, called saccades; stops, called fixations; and movements backward, called regressions (Conklin & Pellicer-Sánchez 2016). During a saccade, or an eye movement, the reader brings a new section of text to attention. During a fixation, or a stopped period, the reader gathers meaning from the text. Regressions, which are less common, can represent a simple correction for an optical overshoot or an attempt to reconcile some kind of perceived error or ambiguity in the meaning of the text. These data points can also be divided among temporal lines, into early and late measures. Early

measures, which can include things like first-fixation length, reflect early, automatic processes. Skipping a word, for instance, is highly automated and reflects the word recognition process, which occurs very early in reading. Late measures, by contrast, reflect later, more conscious and strategic types of processing. In most cases, these can be associated with effort: a word which requires re-reads, multiple fixations, etc. is most likely a difficult word that the reader has to make a more concerted effort to understand and integrate into context.

Eye-tracking for SLA research purposes

Eye-tracking is particularly valuable in SLA research for a number of reasons. To begin with, online data-production methods, like eye-tracking, can be favorably compared to more traditional, or "offline," methods of gathering linguistic data. Chun (2003) notes that SLA research often gives pride of place to production data—e.g., language production and learning outcomes. However, this priority is not necessarily helpful for learning more about a more passive skill, like reading, that does not naturally produce any language artifacts. Off-line methods of gathering information about how a reader processes text require a secondary task in order to provide measurements, and this secondary task can introduce confounding variables: for instance, a think-aloud protocol requires the subject to have certain spoken-language skills that are not relevant to the reading task itself (Conklin & Pellicer-Sánchez, 2016). By engaging in a more naturalistic type of reading task, as is possible with eye-tracking, researchers can avoid the effects of assessment washback.

Overall, while a great deal of information can be gleaned from both performance data and self-reports, they are several steps removed from the actual processes behind those results. Behavioral data, on the other hand, can be more directly reflective of process. The subject creating this data has less control over things like their gaze duration while reading, which can mitigate the effects of learners' assumptions about their own language use. Much as corpus

analysis has changed research into language structure, behavioral data can change research into language processing: there is no reason to depend exclusively on learners' beliefs about what they are trying to do when there is also information about what they really do. Winke, Godfroid & Gass (2013), in an introduction to a special issue of Studies in Second Language Acquisition, concur with the applicability of eye-tracking to SLA research. They emphasize in particular the connection between cognition and acquisition, which has historically been difficult to study through traditional methods. Likewise, Spinnger, Gass, & Behney (2013) suggest that eye-tracking has a great deal of potential for scholars interested in modeling the interactions and connections between language systems in the mind of a bilingual language user.

Winke, Godfroid, & Gass (2013) also emphasize the richness of eye-tracking data. A huge amount of data is also gathered during eye-tracking, with some eye-trackers able to sample up to 2000 times per second. This huge number of samples can be manipulated using software like DataViewer into statistics that represent everything from the length of fixations on articles in a sentence to the total number of eye movements performed while reading an entire passage. In phonological studies, eye-tracking data can even report the length of fixations in a certain letter inside of a word (Alhazmi, Milton & Johnston 2019). This data is also not limited by the researcher's assumptions about which measures will be affected by changing a certain variable: once an eye-tracking data file has been produced, it can be processed repeatedly to report different measurements and even be analyzed to create an entirely new study.

It is also important to note that eye-tracking data can be used to enhance, rather than replace, other methods of data-gathering. Roberts & Siyanova-Chanturia (2013) also exhort the promise of eye-tracking methodology in the field of SLA research. They specifically name the combination of online processing data (like eye-tracking) and metalinguistic data (like post-task reflection) as a rich source of information for SLA researchers. There is no need, these authors

argue, to choose one type of data over the other; rather, they can be used to complement each other, with metalinguistic information guiding the focus of online data analysis. It is increasingly evident that task conditions can affect the manner in which learners process input, and therefore that data-gathering which occurs without requiring a secondary task is enormously important to understanding natural language use.

Known factors in L2 language processing

Eye-tracking has already been used to establish multiple principles of eye-movement in L2 English populations. Some studies, for instance, have confirmed that readers can be affected by the accessibility of grammatical structures in their first language. Bott & Gattnar (2015) examined the reading patterns of Russian students, whose first language uses verb aspect to reflect completeness, and German students, whose first language does not. This study found that German students were less likely to notice conflicting information about completeness, while Russian students were more likely to be surprised and slowed by the same type of information. Sagarra & Ellis (2013) likewise confirmed that adult acquisition of verb morphology is strongly guided by verb patterns in the learner's native language.

It has also been argued that eye-tracking is specifically valuable to researchers concerned with second-language sentence comprehension. As Dussias (2010) writes, eye-tracking has the potential to move SLA syntactic studies away from reported intuitions about grammaticality and correctness, which have been repeatedly proven misrepresentative of actual language use, to better match the more objective methods used in other fields. For syntacticians, who conceptualize processing as a multilayered process of integration and structural decision-making, eye-tracking can reveal the unconscious steps of comprehension. For instance, in the aforementioned study, Sagarra & Ellis (2013) showed that L2 proficiency has significant effects on other reading patterns detected in eye-tracking data. Readers with a lower L2 proficiency read

more slowly, as is to be expected, but also regressed more often and for longer times, suggesting that they specifically have greater difficulties with syntactic integration. SLA research using eye-tracking has great potential to provide information on the processes which English learners use to incorporate grammatical and lexical information in order to understand a single sentence.

Eye-tracking and Arabic-speaking English learners

Eye-tracking research conducted with L1 Arabic speakers does not yet represent a large body of research (Martin, 2011). A few lines of questioning have been well-established, however. One is the matter of "vowel blindness," which has been described in more detail in section 2.1. As previously mentioned, this construct is predicated on the fact that, of Arabic's six vowels, only three are represented by characters. Short vowels in Arabic are represented with diacritics, and while they communicate grammatical information like case and tense, they do not carry content information and can be ignored in favor of context cues (Alsadoon & Heift 2015). It is believed that this results in part or all of the reading difficulties noted in Arabic-speaking English learners (Alhazmi, Milton, & Johnston, 2019): Arabic-speakers transfer their reading strategies from their first to their second language, thereby using a pattern of noticing that is unproductive in English. Eye-tracking methodologies have been used to solidify researchers' understanding of the effects of vowel blindness (Alsadoon & Heift, 2015; Alhazmi, Milton, & Johnston, 2019; Martin, 2011).

Vowel blindness is not, however, the only question about Arabic L1 English reading to which eye-tracking has been applied. Tuninetti, Warren & Tokowicz (2015) used eye-tracking to understand how both Mandarin-speaking and Arabic-speaking English learners process and prioritize syntactic cues. In this study, participants' gazes were tracked while they read a series of sentences, some of which contained syntactic violations of varying types, and participants were also asked to deliver an offline judgment of grammaticality or ungrammaticality to each

sentence. They found that certain violations did prompt more rapid judgments of ungrammaticality, namely noun-article inconsistency: sentences containing a mismatched noun and article pair resulted in measurement differences like longer first-fixation times, which researchers interpreted as greater sensitivity to the mistake; they also prompted more accurate judgments of ungrammaticality. In addition, Tuninetti, Warren, & Tokowicz found that the results of their study did not indicate any evidence of cross-language transfer effects, either from online or offline data sources. This is significant especially given the fact that their participants came from emphatically different language backgrounds, and it raises other questions about whether or not Arabic L1 English learners may display L1-specific effects on the syntactic level as they do on the phonological level.

Anaphor, Binding, and Reflexives

Binding principles

Now let us turn to the specific language issues that this study aims to examine by using eye-tracking: anaphor, binding, and reflexives. So far, eye-tracking has often been used to study pronouns and other referential grammar, due to its ability to report data that suggests conflict recognition and resolution on a syntactic level. For example, Doherty & Conklin (2016) have studied the phenomenon of gender expectations in English reading: namely, that of the singular pronoun “they.” While many English speakers use “they” as a referent to a singular person of unknown gender, and while “they” is increasingly accepted as a chosen personal pronoun, this study shows that “they” does not function smoothly to a similar degree in all contexts. Conflicts between “they” and an antecedent run along two axes: the first, that of gender, and the second, that of number. Doherty and Conklin (2016) used both eye-tracking and participant ratings to determine how disruptions related to gender expectancy, noting that previous experiments had showed readers’ tendencies to search for a plural antecedent when presented with the pronoun

“they” or “them.” Their results showed that nouns with a higher degree of gender-expectancy (for instance, “mechanic,” which is strongly coded male) were processed differently at the earliest stages measured on the eye-tracker than nouns with an encoded gender (for instance, “spokesman). These results seem to show that both expected gender and level of expectancy are accessed very early, much earlier than the syntactic incorporation stages at which an antecedent-referent mismatch would be processed. Participants could easily and quickly incorporate “them” as a referent to a singular noun when the noun did not have a high degree of gender expectancy, challenging the notion that “they” is initially only processed as a plural pronoun: rather, the gender implications of “they” were accessed at least as early as the numerical implications.

One common thread in studies of pronoun reference and resolution is “binding Principle A,” which states that reflexive pronouns “must be linked to a c-commanding noun phrase that matches the reflexive in gender, person, and number” (Chomsky, 1981, as cited in Felser, Sato, & Bertenshaw, 2009). Application of this principle, however, varies from language to language. Some languages do not encode certain factors in reflexive pronouns; others allow for longer or shorter binding distances. Native English speakers have been shown to prioritize this principle in reflexive processing, and only to begin examining discourse cues when binding principle A is compromised by syntactic ambiguity (Nicol & Swinney, 2003; Cunnings, Patterson & Felser, 2014; Cunnings & Sturt, 2014).

Non-native English speakers, on the other hand, can be more flexible in their use of binding principles. Felser, Sato, & Bertenshaw (2009) study application of this principle in Japanese L1 learners of English. This eye-tracking study noted that, although the L1 Japanese speakers had a native-like understanding of principle A, they did not always use it to process reflexive structures. Essentially, although the Japanese L1 participant group knew how to use principle A to understand reflexives in English, they did not always apply this knowledge evenly

during reading. Instead, they were more likely to bind a reflexive to an antecedent outside of the binding range because it was more prominent according to discourse principles.

In a later study, Felser & Cunnings (2012) use eye-tracking methodologies to look at processing of similar reflexive pronoun structures in German L1 English learners. They confirmed the notion that that native speakers applied binding principle A at the first stages of processing, taking the hierarchy of constituents in the sentence into account. The non-native speakers, on the other hand, favored nouns that were more attractive from a discourse perspective, for instance by being mentioned first or mentioned more than once. As a result, they differed from native speakers in choosing incorrect antecedents: questions about whether or not the grammar of the sentence allowed a particular coreferent choice came after the consideration of discourse cues. Together, Felser, Sato & Bertenshaw (2009) and Felser & Cunnings (2012) clearly show a qualitative difference between the processing of pronouns in native and second-language speakers of English, and therefore provide an opportunity for further investigation of these principles in other L1 English learner groups.

Gender and reflexive binding

One of the discourse factors that can give readers pause when choosing a noun to interpret as the coreferent to a pronoun is gender. Cunnings & Sturt (2014) performed a study in which native English speakers were asked to read ambiguous sentences like "John heard about the soldier's picture of himself" and to choose a referent noun for the reflexive pronoun. While participants primarily ended up choosing to interpret the reflexive as referring to the subject of the sentence, because their gaze was tracked while reading, the researchers could note that the participants initially connected the reflexive pronoun to the nearest noun—in the example sentence, "soldier." This shows, firstly, the advantages of a broad online data collection method like eye-tracking: the readers' antecedent choice did not completely represent their actual reading

and processing patterns. Secondly, this study showed the effects of gender conflict: when the second noun did not match the gender of the reflexive pronoun, processing times were longer, despite the fact that the participants would end up identifying the (gender-matched) subject of the sentence as the coreferent.

In an earlier study on native English speakers, Kreiner, Sturt & Garrod (2008) use eye-tracking to directly probe the effects of gender cue conflict on processing patterns. In order to further distinguish the effects of gender cues on English reading, this experiment differentiated between stereotypically and definitionally-gendered nouns—that is, between nouns like "king," which contain gender as part of their lexical definition, and nouns like "surgeon," which are associated with a certain gender as part of a set of world knowledge. Stimuli were presented that contained a noun associated with a particular gender either stereotypically and definitionally. These sentences also contained a gendered reflexive pronoun, which was manipulated to either match or mismatch the gender associated with the referent noun. Two experiments were conducted: the first used anaphoric sentences, in which the noun precedes the pronoun, as in "The priest muttered to himself as he walked," and the second used cataphoric sentences, in which the pronoun precedes the noun, as in "Muttering to himself, the priest walked home." An eye-tracker was used to assess the differences between how these mismatches were resolved. In the anaphoric sentence type, eye-tracking measurements did not show differing processing costs between stereotypical and definitional gender. In cataphoric sentences, the processing costs incurred by a mismatch between a definitionally-gendered noun and its pronoun remained high. However, the processing costs were greatly reduced in cataphoric sentences where the noun was stereotypically gendered, implying that, when gender is provided before the content noun, native English speakers do not access stereotypical gender in the same way they do definitional gender. Kreiner, Sturt & Garrod conclude that there is a qualitative difference between definitional and

stereotypical gender processing. They cite this as evidence for the theory that stereotypical gender information is accessed via a learner's world knowledge, and that it is not stored as part of a word's lexical entry in the way that definitionally-gendered words are. This interplay of knowledge types and conflict resolution poses several intriguing questions for SLA researchers: if accessing and incorporating gender knowledge into a syntactic structure is complex for native English speakers, how much more complex might it be for readers who balance more than one language structure?

Current Studies

It has been established both by instructor experience and experimental evidence that Arabic-speaking learners of English are particularly taxed by the task of reading in English (Alhazmi, Milton, & Johnston, 2019; Alsadoon & Heift, 2015; Clahsen et. al, 2013; Jordan et. al, 2014). The value of eye-tracking as a potential method for investigating these types of phenomenon has also been discussed, both for all language users (Cunnings & Sturt, 2014; Cunnings, Patterson & Felser, 2014; Doherty & Conklin, 2016; Dussias, 2010; Duchowski, 2007; Sturt, 2003) and for English learners in particular (Bott & Gattnar, 2015; Conklin & Pellicer-Sánchez, 2016; Spinner, Gass & Behney, 2013). For the field of Arabic L1 English reading, little research has been done about the effects of L1 Arabic on L2 English syntactic processing (Martin, 2011), and eye-tracking has been discussed as a valuable option for investigating questions about syntactic parsing and integration (Cunnings & Sturt, 2014). The present study will combine knowledge of Arabic-specific reading struggles (Abu-Rabia, 1997) with the framework of Kreiner, Sturt & Garrod's (2008) investigation into gender cue conflict resolution, and by doing so, will begin to address the lack of understanding of syntactic processing in Arabic L1 English readers.

Research Questions

Guided by the research and theories reviewed above, this study examines the following research questions: How do Arabic-speaking learners of English react to and resolve gender cue conflicts between a reflexive pronoun and the noun that it is bound to? How do these processes differ from those of native English speakers?

.

CHAPTER 3: METHODOLOGY

In this chapter, the methodology of the present study is described. The characteristics of the participant groups are reported first, followed by a description of the development and product of the experimental materials. Technological specifications for the eye-tracking set-up are also provided. Finally, the procedures followed during the experiment are outlined in detail.

Participants

15 non-native speakers of English (NNS) participated. All were male. All spoke Arabic as a first language. As a requirement for university admission, all had either met or exceeded a score of 79 on the TOEFL iBT, 550 on the paper-based TOEFL, 6.0 on IELTS, 54 on Pearson's Test of English, or 51 on Duolingo English Test. None reported fluency in a language other than English and Arabic. None reported residency in an English-speaking country for more than one year prior to beginning university study. Average age of NNS participants was 19.75 years (19-21, SD = 0.6 years). Countries of origin were Saudi Arabia, Kuwait, Oman, and the United Arab Emirates. 50% reported five or more years of English language study; 16% each reported 2-5 years, 1-2 years, and 6 months to one year of study.

5 native speakers of English (NS) participated. All were female. None reported living outside of the United States. All spoke English as a first language, and none reported fluency in another language. Average age of NS participants was 18.2 years (18-19, SD = 0.4 years).

All participants were undergraduates at the University of Alabama. All were recruited from their required first-year composition courses (EN101 or EN102 for NS participants; EN120 or EN121 for NNS participants) via researcher visits to their classrooms and email follow-up. All

had normal or corrected-to-normal vision. All were compensated for their participation with extra credit in their composition course.

Materials

Experimental materials were derived from those used by Kreiner, Sturt & Garrod (2008), the noun list produced by Kennison & Trofe (2003), and the Corpus of Contemporary American English (Davies). Experimental sentence structures were derived from the experimental sentences used in Kreiner, Sturt & Garrod's (2008) procedure. The stereotypically-gendered nouns which were paired with reflexive pronouns in the experimental sentences were selected from a list of gender-associated nouns by Kennison & Trofe (2003), then cross-referenced with the top 5,000 most common words found in COCA in order to enable better comprehension for NNS participants. The nouns chosen for the study were only those that were at least one standard deviation outside of the "gender-neutral" range established in Kennison & Trofe (2003) and could also be found within the list of the top 5,000 words in COCA. In order to minimize frequency effects, definitionally-gendered nouns were selected based on their proximity to the stereotypically-gendered nouns on the frequency-ordered list so that rarer nouns of each type were paired together. All other words in the experimental and filler sentences were found within the top 2,000 words in COCA.

The experimental procedure involved 67 sentences total, each followed by a yes-no comprehension question. A practice block of three sentences, each involving an anaphoric reflexive, was shown first. After that, participants were guided through four experimental blocks of 16 sentences each (64 in total). Of those 64, 32 were experimental and 32 were filler sentences derived from Kreiner, Sturt & Garrod's (2008) filler materials; these were interleaved randomly throughout each block. Of the 32 experimental sentences, each participant saw 8 sentences in

each condition: stereotypical/match, stereotypical/mismatch, definitional/match, and definitional/mismatch.

Table 1. ***Four conditions.***

	Definitional	Stereotypical
Match	Today the *lady* turned off the computer before introducing *herself* to the new worker.	Today the *nurse* turned off the computer before introducing *herself* to the new worker.
Mismatch	Today the *lady* turned off the computer before introducing *himself* to the new worker.	Today the *nurse* turned off the computer before introducing *himself* to the new worker.

Each stereotypical/definitional noun pair was embedded into two different sentences so that every participant saw each noun once. The average length of an experimental sentence was 12.28 words; average filler sentence length was 12.68 words.

The experimental items were divided up into four counter-balanced lists so that, across participants, each noun was presented in each of the four conditions, but individual participants saw each noun only once. The same filler items were used for each list. Once generated, these lists were automatically randomized by the Experiment Builder software in order to reduce the effect of item-order on experimental results.

Materials were formatted to Spinner, Gass & Behney's (2013) recommendations for ecological validity. To match the majority of reading materials used in U.S. academic settings, sentences were presented left-justified, in black Times New Roman font, and against a white backdrop. However, due to the distance requirements for the eye-tracking camera, they were presented in 18 point font to prevent eye strain. All sentences were presented as one complete line of text with margins of one inch from all sides of the display screen.

A visualization of the sentence format and the eye movements recorded in each condition can be seen in Figures 1-4.

Figure 1. ***Definitional match condition.***

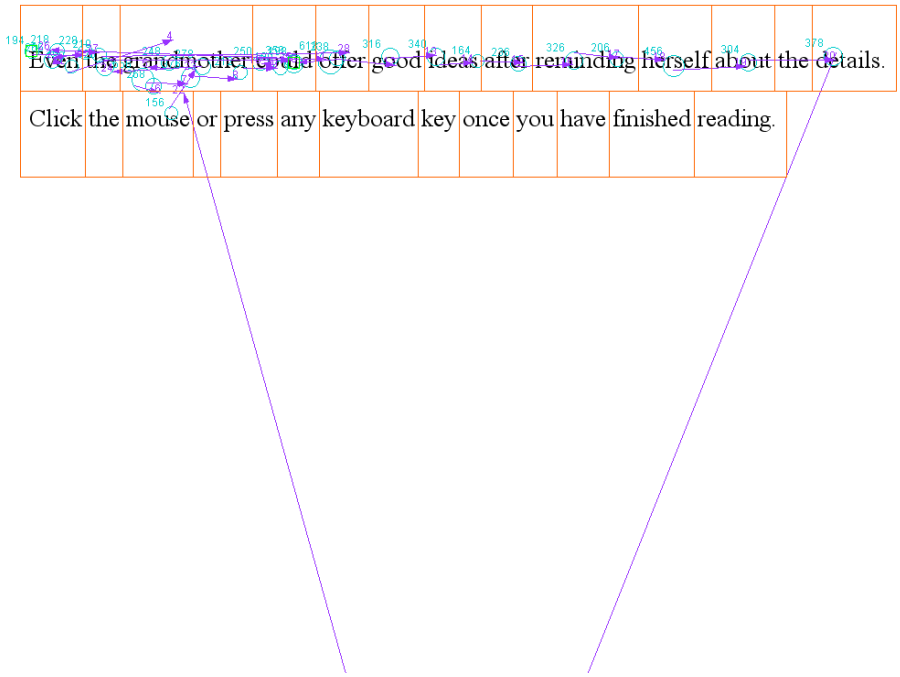


Figure 2. ***Definitional mismatch condition.***

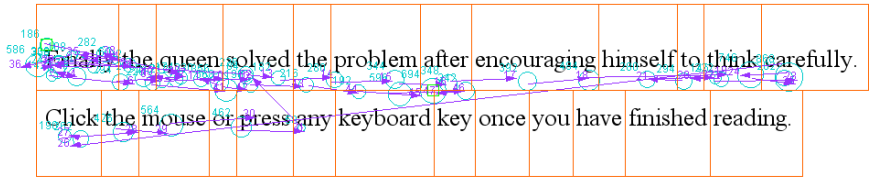


Figure 3. ***Stereotypical match condition.***

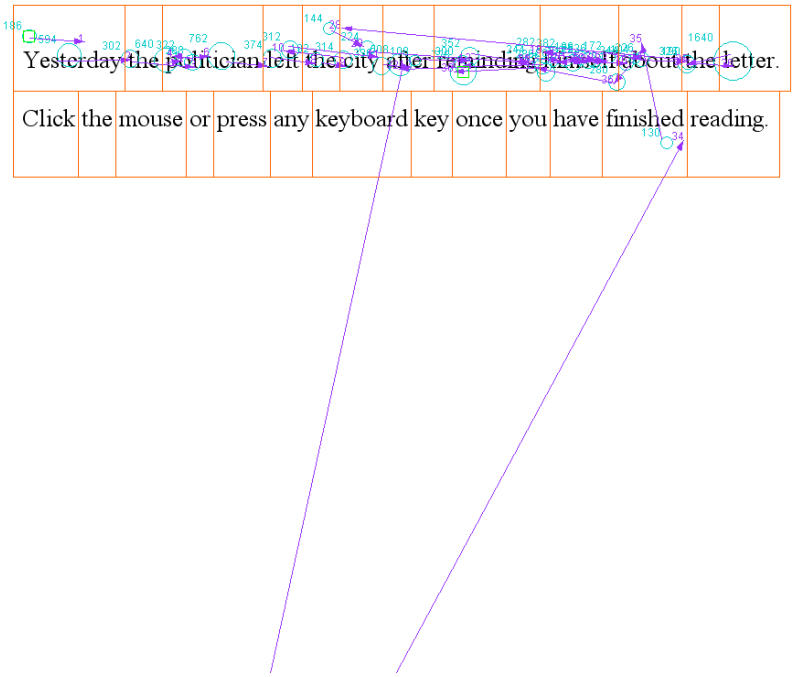
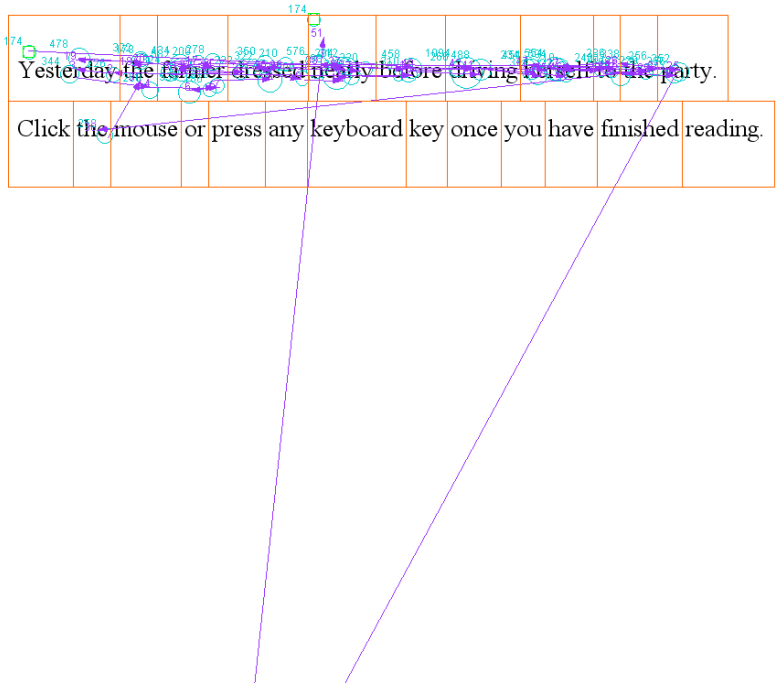


Figure 4. ***Stereotypical mismatch condition.***



Apparatus

The experiment was conducted using an SR Research EyeLink Portable Duo mounted on a laptop. The front edge of the laptop was 23cm from the edge of the desk and its screen was

roughly 45cm from the position of the participants' eyes. A stationary chin rest was used in order to stabilize participants' heads for more precise recording. Data was recorded at a sampling right of 500 Hz. Viewing was binocular, but recording was only taken of the right eye.

Procedure

Upon arrival to the lab, participants were greeted by the researcher and provided with a paper copy of the consent form, which they had previously seen in digital form. The researcher answered any questions before the consent form was signed. They were also asked for their first and last name and the name of their composition instructor in order to report earned class credit; all other study information was recorded under a participant number to preserve anonymity.

Then, participants were seated at the eye-tracking desk and informed verbally that the study was comprised of three sections: a paper survey about their general academic and linguistic background (Appendix A), a sequence of sentences and questions on the display computer (Appendix B) during which gaze data would be collected, and a post-test recall task (Appendix C), also on paper. Once participants completed the survey, they were asked to begin reading from the laptop. The experiment began with a calibration sequence in which participants were asked to focus on a series of dots on the screen in order to test position accuracy. In addition, a drift check was performed before each trial in which the computer displayed a blank screen with one dot positioned where the first word of the trial sentence would appear. When participants fixated on that dot, the trial was manually triggered.

The experiment began with one practice block of three sentences, each of which contained a gender-matched anaphoric reflexive pronoun, and their respective comprehension questions. Participants used either the keyboard or the mouse to move from the sentence screen to the question screen. Participants were asked to answer the comprehension question by pressing the "Z" key to answer "yes" and the "M" key to answer "no." Instructions to this effect

were presented on each screen. In order to obscure the focus of the experiment, comprehension questions for experimental items did not directly concern the referent of the anaphor. Data from 91% of experimental trials were used for analysis; the remaining 9% were rendered unusable due to camera or recording failure. Participants answered the comprehension question correctly for 86% of usable trials. The eye-tracking section of the experiment was completed in a period of 20-35 minutes for each participant.

After the recording period of the experiment, participants were asked to complete the post-test recall task (Appendix C), which provided additional motivation for comprehension and served as a distractor from the main purpose of the study. The word bank on the post-test recall task contains 88 possible nouns, of which 32 were experimental nouns from the eye-tracking procedure. The rest were taken from Kennison & Trofe (2003) and were not present in the experimental materials or filler content. Participants were asked to mark any and all words from the word bank which they remembered seeing during the experiment.

CHAPTER 4: RESULTS

This chapter provides a summary and analysis of the data derived from the present study. First, the procedures for creating the data report are described. Then, the data itself is reported, both in plain numerical form and then in longform text which indicates the basic implications of a variety of measures. This longform data reporting is divided by participant group (first NS statistics, then NNS statistics) and then by region of interest.

Data Analysis Procedure

Data was first visualized and compiled using DataViewer and then analyzed using R. For analysis, three interest areas were isolated from each sentence:

Table 2. ***Interest areas.***

Region	*Example text*	*For the sentence "Recently the hunter worked alone after reminding himself about the competition."*
Antecedent	Hunter	
Reflexive	Himself	
Spillover	About	

These are the primary regions in which processing delays can be expected to be shown. Antecedent and reflexive sections draw more of the reader's attention because, if the sentence does give conflicting cues, they will be found in these two items. The spillover section was included because processing of a more complex item can often occur parafoveally (in peripheral vision, while the gaze is focused on the subsequent section) causing longer fixation times for the word following the complex item.

In addition, four types of gaze data are reported. *First-fixation* duration, which reflects the length of the reader's first stationary focus on the area, is assumed to be connected to early reading processes, especially lexical access. The *first-pass* reading time, which is also reflective of earlier stages of meaning extraction, is calculated from the sum of all fixations a reader's gaze makes in a region before exiting to either the right or the left. *Regression-path* duration similarly begins when a reader's eyes first enter a region, but only ends when the reader's gaze moves to a subsequent section of the sentence and is therefore different from first-pass reading time when a reader returns to an earlier section of the sentence before finishing the sentence. Finally, second-pass duration begins the second time a reader's gaze enters a particular area, and is therefore connected with later stages of processing, such as resolving ambiguities and other forms of syntactic integration.

Reporting

Table 3 displays means, standard deviations, and analyses of variance (ANOVAs) for each measurement in each relevant area for the NS participant group (matching data from the NNS participants can be seen in Table 4). The ANOVA function followed a 2x2 design—that is, it calculated the strength of association between both gender variables (stereotypical/definitional and match/mismatch) and the length of the measurement. Marginal significance is shown by a # symbol.

NS participant results will be presented first to serve as a baseline of expected fluent reader behavior. The results of the study will be organized first by region, then by different measures reported within that region, if any. NNS participants will be presented second, and the same statistics will be reported. General comparisons between the two groups will be observed in the NNS section, and detailed interpretation of this comparison will be presented in the Discussion.

Table 3. ***NS Results***

Region	Reflexive				Spillover	Antecedent
Measure Mean durations in ms	First-fixation Mean *(SD)*	First-pass Mean *(SD)*	Regression-path Mean *(SD)*	Second-pass Mean *(SD)*	First-fixation Mean *(SD)*	Second-pass Mean *(SD)*
Condition Definitional-match	198.2857 (107.1304)	198.2857 (107.1304)	220.8571 (122.3042)	204.2857 (60.22656)	195.3333 (55.22077)	210.8667 (82.60134)
Definitional-mismatch	289.4118 (111.3031)	315.6471 (137.4104)	360.9412 (275.9023)	200 (120.3614)	264 (113.5782)	275 (124.5338)
Stereotypical-match	249.7 (80.06649)	280.9 (126.9007)	324.2 (220.7803)	149 (55.933)	335 (245.9245)	254.6667 (86.98659)
Stereotypical-mismatch	247.2222 (77.35674)	257.2222 (68.00048)	280.1667 (131.1785)	249.25 (144.3644)	178.8 (87.55113)	216.3333 (146.1226)
*ANOVA*s	F;P	F;P	F;P	F;P	F;P	F;P
Noun Matching	3.914; 0.0517#	0.058; 0.811	1.103; 0.297	0.079; 0.780	2.135; 0.150	0.851; 0.429
Noun Type	0.138; 0.7111	0.123; 0.727	0.165; 0.686	0.561; 0.458	0.081; 0.777	0.701; 0.404

Native speaker reflexive region

First-fixation

NS participants' average first-fixations in the reflexive region ranged between 198 ms (definitional-match condition) and 289 ms (definitional-mismatch condition). The ANOVA test found a marginally significant relationship ($p = 0.0517$) between noun matching and first-fixation times in this region: in sentences where the gender of the noun and the gender of the reflexive pronoun did not match, NS participants' first fixation on the reflexive pronoun was generally longer than in sentences where the gender cues matched. The ANOVA did not show a significant relationship between first-fixation time in this region and noun type; both definitionally-gendered and stereotypically-gendered nouns appeared to produce initial fixations of a similar length.

First-pass

NS participants' average first-pass times in the reflexive region ranged between 198 ms (definitional-match condition) and 315 ms (definitional-mismatch condition). The ANOVA test found no significant relationship between either noun matching or noun type variables and first-pass times in this region; all noun conditions appeared to produce initial pass times of a similar length. In addition, first-pass times in the definitional-match condition were exactly the same as first-fixation times: none of the NS participants made a second fixation in the reflexive region in a sentence where the gender of the noun and reflexive matched.

Regression-path

NS participants' average regression-path times in the reflexive region ranged between 220 ms (definitional-match condition) and 360 ms (definitional-mismatch condition). The ANOVA test did not find a significant relationship between either noun matching or noun type variables and regression-path times in this region: noun conditions did not appear to affect the

length of time a participant spent on the reflexive region before moving on to the rest of the sentence.

Second-pass

NS participants' average second-pass times in the reflexive region ranged between 149 ms (stereotypical-match condition) and 249 ms (stereotypical-mismatch condition). The ANOVA test found no significant relationship between either noun matching or noun type variables and first-pass times in this region; in all noun conditions, participants spent a similar amount of time re-reading the reflexive pronoun.

Native speaker spillover region

NS participants' average first-fixation times in the spillover region ranged between 178 ms (stereotypical-mismatch condition) and 335 ms (stereotypical-match condition). The ANOVA test found no significant relationship between either noun matching or noun type variables and first-fixation times in this region; all noun conditions appeared to produce first fixations of a similar length.

Native speaker antecedent region

NS participants' average second-pass times in the antecedent region ranged between 210 ms (definitional-match condition) and 275 ms (definitional-mismatch condition). The ANOVA test found no significant relationship between either noun matching or noun type variables and first-pass times in this region; in all noun conditions, participants spent a similar amount of time re-reading the referent noun.

Table 4. ***NNS Results***

Region	Reflexive				Spillover	Antecedent
Measure Mean durations in ms	First-fixation Mean *(SD)*	First-pass Mean *(SD)*	Regression-path Mean *(SD)*	Second-pass Mean *(SD)*	First-fixation Mean *(SD)*	Second-pass Mean *(SD)*
Condition						
Definitional-match	299.625 (139.4972)	428.9375 (139.4972)	566.4375 (461.797)	336 (303.9618)	234.1818 (81.55221)	481.2593 (356.8832)
Definitional-mismatch	258.6667 (84.17747)	329.1111 (144.8971)	628.7407 (727.4893)	418.6667 (430.8363)	265.2308 (48.95943)	472.0345 (377.0341)
Stereotypical-match	248 (71.03118)	305.7931 (105.4641)	422.7586 (344.5819)	374.8333 (410.72)	234.5882 (52.02771)	440.32 (364.0945)
Stereotypical-mismatch	305.9655 (118.7538)	429.8966 (335.6938)	1039 (1957.412)	228.9091 (63.91472)	254.6154 (76.5436)	382.9062 (304.8011)
*ANOVA*s	F;P	F;P	F;P	F;P	F;P	F;P
Noun Matching	0.156; 0.694	0.058; 0.811	2.941; 0,0891#	0.079; 0.780	2.135; 0.150	0.851; 0.429
Noun Type	0.043; 0.836	0.123; 0.727	0.370; 0.5445	0.561; 0.458	0.081; 0.777	0.701;0.404

Reflexive region

First-fixation

NNS participants' average first-fixations in the reflexive region ranged between 248 ms (stereotypical-match condition) and 305 ms (stereotypical-mismatch condition). Unlike in the case of the NS participants, the ANOVA test did not find a relationship between noun matching and first-fixation times in this region: NNS' first fixations on the reflexive pronoun did not appear to be longer when the gender of the noun and the gender of the pronoun did not match. The ANOVA also did not show a significant relationship between first-fixation time in this region and noun type; both definitionally-gendered and stereotypically-gendered nouns appeared to produce initial fixations of a similar length.

First-pass

NNS participants' average first-pass times in the reflexive region ranged between 305 ms (stereotypical-match condition) and 429 ms (stereotypical-mismatch condition). The ANOVA test found no significant relationship between either noun matching or noun type variables and first-pass times in this region; all noun conditions appeared to produce initial pass times of a similar length. Unlike NS participants, first-pass times in the definitional-match condition were not the same as first-fixation times: NNS participants consistently made more than one fixation in this region.

Regression-path

NNS participants' average regression-path times in the reflexive region ranged between 422 ms (stereotypical-match condition) and 1039 ms (stereotypical-mismatch condition). Unlike in the NS participant group, the ANOVA test did find a marginally significant relationship ($p = 0.0891$) between noun matching and regression-path times in this region: in sentences where the gender of the noun and the gender of the reflexive pronoun did not match, NS participants took

longer to move on from the reflexive region to following parts of the sentence. The ANOVA did not show a significant relationship between regression-path time in this region and noun type; both definitionally-gendered and stereotypically-gendered nouns appeared to produce regression-path times of a similar length.

Second-pass

NNS participants' average second-pass times in the reflexive region ranged between 228 ms (stereotypical-mismatch condition) and 418 ms (definitional-mismatch condition). The ANOVA test found no significant relationship between either noun matching or noun type variables and first-pass times in this region; in all noun conditions, participants spent a similar amount of time re-reading the reflexive pronoun.

Spillover region

NNS participants' average first-fixation times in the spillover region ranged between 234 ms (definitional-match condition) and 265 ms (definitional-mismatch condition). The ANOVA test found no significant relationship between either noun matching or noun type variables and first-fixation times in this region; all noun conditions appeared to produce first fixations of a similar length.

Antecedent region

NNS participants' average second-pass times in the antecedent region ranged between 382 ms (definitional-match condition) and 481 ms (definitional-match condition). The ANOVA test found no significant relationship between either noun matching or noun type variables and first-pass times in this region; in all noun conditions, participants spent a similar amount of time re-reading the referent noun.

CHAPTER 5: DISCUSSION

This chapter discusses the results of the experiment. First, it summarizes the basis for and goals of the data analysis. Secondly, it provides interpretation of the results presented in Chapter 4 for each measurement reported. This begins with measurements which did not appear to be affected by the experimental conditions, and it concludes by describing and interpreting the measures that were most affected by the experimental conditions for both NS and NNS groups. Potential explanations for the reported observations are offered. Finally, this chapter presents the implications of these data for both research and ESL pedagogy.

General Discussion

Native Arabic speaking students of English have frequently been observed to struggle with reading more than students from other L1 backgrounds. The root of this difficulty has been suggested to be related to the contrast between the Arabic and English alphabets, the differences between their writing systems, and other word-level or phoneme-level concerns, especially "vowel blindness." As a result, research on Arabic L1 English reading has primarily focused on these areas, and researchers have used eye-tracking techniques to provide convincing evidence that Arabic L1 students reading English are affected by patterns of letter-recognition that do not match the habits of native English readers. On the level of single words, the differences between Arabic L1 and English L1 reader processing are clear; however, similar information about sentence-level processing is sparse. Because it is unlikely that the differences do not continue to the level of syntactic processing, this study aimed to address the knowledge gap about Arabic English learners' syntactic processing during reading.

The current study followed an established design by Kreiner, Sturt & Garrod (2008) in order to test the processing costs incurred on Arabic L1 and English L1 readers when a sentence provides mismatching gender cues. Nouns that are gendered by either definition or by stereotype were embedded in sentences containing anaphoric reflexive pronouns, and these pronouns were manipulated to either match or mismatch the gender of the preceding noun. Eye-tracking data was collected while participants read and responded to questions about these sentences. Six different measurements (reflexive first-fixation, first-pass, regression-path, and second-pass times; spillover first-fixation time; and antecedent second-pass time) were statistically analyzed to determine the effects of gender type and mismatch in each participant group. This provided a picture of the processing costs involved in reading sentences with mismatched gender cues. Then, results were compared between participant groups in order to determine the differences in processing cost between Arabic L1 and English L1 participants.

Between the two groups, NNS data showed longer total dwell times across conditions. An independent sample t-test between NS and NNS groups' mean dwell times found a statistically significant difference in both the noun and reflexive pronoun interest areas (noun: $t=8.54$, $p=1.795e-15$; reflexive: $t=4.3$, $p=2.9e-5$). This is consistent with expectations about the learners' level of experience in English, as well as with research establishing Arabic L1 English learners' difficulty with reading. NS average fixation times were consistent with expectations about reading speed for proficient English L1 readers. This suggests that the data from the current study is not outside of usual reading norms.

The results for both groups did not show correlation between any eye-tracking measure and the gender type (stereotypical and definitional) of the noun in the experimental sentence. This is consistent with Kreiner, Sturt & Garrod's (2008) results with native English speakers, where differences between definitional and stereotypical gender resolution did not appear in

anaphoric sentences. This also, however, suggests that the NS and NNS groups had a similar understanding of the stereotypical nouns' gender associations. If an NNS did not, for instance, associate "nurse" with a female gender, we would expect to see less of a processing cost when "nurse" was paired with the male reflexive "himself." This did not occur, suggesting that the NNS did have established perceptions of stereotypical gender associations that were similar to the NS group.

There was also no statistically significant relationship for either group between the mismatch variable and the first-pass reflexive measure, the second-pass reflexive measure, the first-fixation spillover measure, or the second-pass antecedent measure. This was likewise similar to Kreiner, Sturt & Garrod's (2008) results with native English speakers. Firstly, this indicates that parafoveal processing in this type of reading study is not likely to be a concern. Perhaps because the content of the reflexive pronoun is a reference and a gender, not a self-contained noun symbol, learners are less likely to move on to the next word while they are still processing the reflexive. Instead, their gaze may be more likely to return to previous parts of the sentence while they continue to process the implications of the reflexive. Secondly, the lack of correlation between second-pass reading time for the antecedent—the subject of the sentence—and the mismatch variable could be rooted in two things. While it is possible that a mismatched noun does not cause readers to return to the antecedent for a longer time, it seems more likely that a matched noun does not receive a second reading pass at all. Because DataViewer reports second-pass reading time for an interest area that does not receive a second pass as a missing value, the means for this measurement do not represent sentences which learners process in a completely linear fashion. While both first-pass and second-pass times were not correlated with either noun variable within either participant group, it is unlikely that this is a result of similar syntactic processing; this will be addressed in following paragraphs.

Among NS participants, only one eye-tracking measure was statistically associated with either of the gender variables. The length of the first-fixation time in the reflexive region was marginally correlated to the mismatch variable: that is, in a sentence where the gender of the noun did not match the gender of the pronoun, native English-speaking participants were more likely to fixate on the reflexive pronoun longer the first time they looked at it. This measure is believed to represent the earliest trackable reading processes; it reflects an implicit action on the part of the participant and is connected to lexical recognition and retrieval. This indicates that native English-speakers are affected by gender cue mismatch at the very beginning of processing. In this case, they appear to immediately access the meaning of a word and understand that it will be difficult to syntactically integrate into the rest of the sentence.

However, this effect appears to be short-lived. The initial fixation on a mismatched pronoun is longer, but the first pass on that word—the combination of all the fixations that occur before the participant's gaze moves to another section—is not. Regression-path time in the reflexive region and second-pass time in both the reflexive and antecedent regions are likewise unaffected by the match variable, meaning that the NS participants' early detection of the mismatch co-occurred with early resolution. A gender mismatch did not cause NS participants to need more fixations on the word, more time reading the antecedent a second time, or more time returning to the pronoun after finishing the sentence, thereby indicating that the mismatch effect is isolated to the earliest of reading processes. Moreover, because second-pass reading times are believed to be connected to later reading processes, this result indicates that NS participants are able to resolve gender conflict before syntactic integration even fully occurs.

NNS participant data, on the other hand, did not show a statistical correlation between first-fixation time and the mismatch variable. Instead, the only measure that appeared to be affected by a gender cue mismatch was regression-path reading time, which measures the length

of time between when a participant's gaze enters a region for the first time and when it moves from that region to one later in the sentence. This means that regression-path time includes the time during which a reader regresses from, for example, a reflexive pronoun to its antecedent noun.

While this measure includes first-fixation time, NNS participants did not have longer first-fixation times in sentences which included a gender cue mismatch. This indicates that, unlike NS participants, they did not recognize the conflict between the reflexive and its antecedent immediately. Rather, the conflict was discovered during subsequent fixations and regressions to earlier parts of the sentence. This indicates that gender cue knowledge is triggered later in the NNS reading process, and not simply because NNS reading times were longer across the board. Lexical retrieval did, indeed, take longer for NNS participants, as reflected by their longer first-fixation times, but the gender cue mismatch cost was incurred after lexical retrieval.

This indicates a qualitative difference between the processes by which English L1 and Arabic L1 readers resolve a gender cue mismatch. While the native English-speakers can identify and resolve the mismatch at earlier stages associated with lexical retrieval, Arabic-speaking English learners do not appear to recognize the mismatch until stages of processing associated with the move from lexical recognition to syntactic integration. Both groups, however, appear to resolve the conflict fairly quickly once it is recognized: mismatch did not affect NNS' measures that reflect later processing, like second-pass reading time. Several explanations for this difference are possible. Firstly, the Arabic L1 readers may find gender to be a less relevant aspect of a pronoun, and therefore do not retrieve it immediately upon accessing the meaning of "himself" or "herself." However, the presence of gendered third-person pronouns in Arabic makes this irrelevance somewhat unlikely. Secondly, the Arabic L1 reader may not be able to hold the features of the referent noun in working memory as well as the English L1 reader: while

a native speaker can retain an impression of the subject of the sentence as they process the words in between the noun and pronoun, the cognitive load of the Arabic L1 reader is heavier, and therefore they are more likely to have to return to the subject of the sentence in order to comprehend a referential pronoun.

This difference between mismatching cues' effects on NS and NNS readers may also have to do with non-native speakers' looser adherence to binding principle A. As proposed by Felser, Sato, & Bertenshaw (2009) and Felser & Cunnings (2012), native English speakers apply this binding principle at very early stages of reading processing, providing a stronger link between coreferents, which would therefore cause more of an immediate interruption if compromised. However, non-native English speakers were found to rely on other markers in written discourse to establish coreference, a process similar to the problem-solving that native English speakers tend to use only when they come across ambiguity or conflict. If this is the case, these results suggest that these Arabic-speaking English learners perform syntactic processing in a similar order to English learners from other language backgrounds.

On a larger scale, these results suggest that L1 Arabic and L1 English speakers use different processes to read English, not just the same processes at varying speeds. While studies have shown this type of difference in visual parsing of words and letters, this study indicates that the differences between these groups extends into higher-order reading processes of syntactic binding as well. In addition, given corresponding outcomes in studies like Tuninetti, Warren & Tokowicz (2015), these results suggest that, as Arabic L1 students' English proficiency grows, their reading patterns may become more similar to students from other language groups.

Research Implications

Further research can clarify the effects of a variety of factors on Arabic L1 English learners' syntactic processing. Firstly, Arabic learners of English can be compared, using eye-

tracking, to groups of English learners from other language backgrounds performing similar tasks. Felser, Sato & Bertenshaw (2009) showed that Japanese L1 English learners were likely to rely on discourse cues over English binding principles; it is not yet clear if this type of prioritization had an effect on the NNS group in the current study. By presenting other L1 groups with sentences that contained gender cue conflict, or by presenting Arabic-speaking English learners with stimuli that tested binding principles in situations with more than one possible referent, research could clarify how the two constructs interact. It may be that Arabic L1 English learners' response to gender cue conflict is similar to the response of learners from other L1 backgrounds; it may also be that Arabic L1 English learners use the same coreferent binding strategies to resolve gender cue conflict as they do to resolve other types of coreferent ambiguity.

Research can also clarify the effects of proficiency across Arabic L1 English learners. Conducting this experimental protocol on participants grouped into low, medium, and high-proficiency bands, perhaps according standardized test scores, could reveal how Arabic L1 students adjust their reading patterns as they become able to retrieve lexical information more quickly or retain noun information in their working memory more easily. Learners at a higher level of proficiency read more quickly, but this could occur because their reading strategies become more native-like or because their qualitatively different processes become more efficient. By incorporating proficiency differences, either through a longitudinal or a cross-sectional design, researchers could discover how Arabic L1 English learners' reading changes over time.

Finally, further research can be used to clarify the role and differentiation between world knowledge and lexical knowledge. If, as Kreiner, Sturt & Garrod suggest (2008) suggest, stereotypical and definitional gender associations are stored differently in native English speakers, their second experiment could be conducted using Arabic L1 participants. In native

English speakers, stereotypical gender cue conflict processing costs were offset when the pronoun, and therefore the gender information, came before the noun. The same was not true for definitionally-gendered nouns. By observing the differences between how Arabic L1 students resolve gender cue conflict in anaphoric (pronoun following noun) and in cataphoric (noun following pronoun) conditions, researchers could glean more information about how second-language learners store gender associations and how world knowledge is connected to second-language frameworks of understanding.

Pedagogical Implications

It is difficult to draw a direct line from the results of this study to a particular pedagogical method or strategy. However, pedagogy should certainly evolve in response to the reading needs of Arabic L1 English learners, especially at institutions like the University of Alabama that host hundreds of them.

Tentative possibilities for pedagogical adjustment include working memory training. If the delayed processing cost incurred by Arabic L1 learners in this experiment was, in fact, a result of overburdened working memory, Arabic-speaking English students could be trained using a self-paced reading task that displays only one word at a time. This would call attention to the regressions made as they read sentences containing coreferent nouns and pronouns, encouraging more conscious noting of the factors, like number and gender, that mark the subject of the sentence as connected to its coreferent pronoun, thereby reducing the need for costly regressions.

As previously mentioned, eye-tracking studies with Arabic learners of English have provided convincing evidence that phoneme recognition and subsequent lexical identification are strongly connected to Arabic-speaking learners' difficulty with reading (Martin, 2011). If the nature of Arabic L1 students' phonemic awareness is related to the processing costs observed in

this study—for instance, if Arabic L1 learners are less likely to differentiate “himself” and “herself” because the difference is comprised partially of vowel contrasts—additional phonetic training could be beneficial for Arabic-speaking English learners. By providing more intensive instruction in pronunciation and letter recognition, ESL teachers can raise students’ awareness of phonetic changes and their centrality to reading comprehension.

CHAPTER 6: CONCLUSION

This chapter synthesizes the purposes and outcomes of the entire experiment. First, it provides a summary of the background, procedures, and results of the experiment. Then, it discusses the limitations of the experiment. Finally, it describes possibilities for future research on the basis of the results of this study.

Summary

While research on native Arabic-speakers' English reading processes has previously made use of eye-tracking methods, these investigations have almost always been focused on the phonological and morphological levels. There is convincing evidence that the differences between Arabic and English alphabets, orthographies, and writing systems are partially responsible for the difficulty that native Arabic speakers have when reading English. There is no reason to believe, however, that the influence between language systems ends at the word level. This study begins to investigate the differences in syntactic processing during reading between Arabic L1 and English L1 groups.

This study used an eye-tracking protocol, along with a language survey and a post-test recall task, to gauge how non-native English speakers (from Arabic backgrounds) and native English speakers responded to gender cue conflicts in written sentences. Two sets of nouns, one set gendered by definition and the other by stereotype, were selected and embedded into sentences that incorporated a reflexive pronoun anaphor. Based on eye-movement measurements taken while participants read sentences in four different conditions—definitional/matching, definitional/mismatching, stereotypical/matching, and stereotypical/mismatching—this study compared how NS and NNS participants reacted to conflicting gender information.

Data analysis indicated that NS and NNS reacted to the syntactic conflict differently. While NS appeared to recognize and resolve the conflict very early in the reading process, NNS recognized and resolved the conflict somewhat later. This suggests that there are qualitative differences between the ways that English L1 and Arabic L1 readers access lexical information, integrate gender into a sentence structure, and bind coreferent elements together. Implications of this information for research and pedagogy were discussed.

Limitations

There are several limitations to this study. The first is the size of each group and the difference between the two. Although the ANOVA test accounts somewhat for the size of a pool of data, and therefore the statistical significance of the measures analyzed in this experiment hold for this group, stronger associations and more generalizable results would be derived from a larger group of participants, particularly of English L1 participants—preferably at least fifteen, although similar studies have also been performed with participant numbers in the 90s (Kreiner, Sturt & Garrod, 2008).

A second limitation to this study pertains to gender imbalance between NS and NNS participant groups: all NNS participants were male, and all NS participants were female. Although previous studies of L1 English speakers did not find a statistically significant difference between the strength of gender association that men and women assigned to the nouns used in this study (Kennison & Trofe, 2003), this may not hold true across languages. Stronger conclusions could be made from this study if both groups were closer to half male and half female. Due to the makeup of the Arabic-speaking student group at the research university, this was not possible during the present study.

Finally, the stimuli (both noun and sentence structures) were not normed on Arabic L1 populations. They had been normed previously by Kreiner, Sturt & Garrod (2008) and Kennison & Trofe (2003), but it is possible that these sentences and questions presented a particular difficulty for Arabic L1 participants.

Future research

Based on this study, future SLA research can investigate the localization of these effects: they may be limited to Arabic L1 English learners, but they may also be applicable to all English learners, regardless of language background. These processing differences may only be present in sentences which contain a gendered anaphoric reflexive pronoun; they may also be elicited in tasks that involve binding and coreference without gender cues. Finally, these processing differences may occur primarily in Arabic L1 English learners at a certain level, and future studies can establish whether or not these effects are reduced or disappear as Arabic L1 learners become more proficient in English.

Future research can also investigate specialized reading instruction for Arabic-speaking students. If Arabic students struggle to retrieve all relevant lexical information when they encountered a word, or if their syntactic processing differs from native English speakers because they do not recognize critical phonemes as readily, focused instruction in these areas may increase reading speed and comprehension.

Finally, further research can investigate whether or not these reading process differences are a disadvantage at all. It is irresponsible to assume that a process which is not native-like is inherently problematic. As an Arabic L1 English learner becomes more proficient in the language, they may find that their reading speed and comprehension level is better if they are not taught to read like native speakers, but rather assisted in developing their unique reading strategies. Any pedagogy developed to assist Arabic L1 English learners in reading should be

carefully tailored to their actual language needs and not simply the differences between their process and a native English speaker's.

REFERENCES

Abu-Rabia, S. (1997). Reading in Arabic orthography: The effect of vowels and context on reading accuracy of poor and skilled native Arabic readers in reading paragraphs, sentences, and isolated words. Journal of Psycholinguistic Research, 26, 465-482.

Alhazmi, K., Milton, J. & Johnston, S. (2019). Examining 'vowel blindness' among native Arabic speakers reading English words from the perspective of eye-tracking. *System*, *80*, 235-245.

Alsadoon, R. & Heift, T. (2015). Textual Input Enhancement for Vowel Blindness: A Study with Arabic ESL Learners. *The Modern Language Journal, 99*(1), 57-79.

Bott, O. & Gattnar, A. (2014). The cross-linguistic processing of aspect – an eyetracking study on the time-course of aspectual interpretation in Russian and German. *Language, Cognition, and Neuroscience, 30*(7), 877-898.

Chun, D. (2013). Contributions of Tracking User Behavior to SLA Research. *CALICO Journal*, *30*, 256-262.

Clahsen, H., Balkhair, L., Schutter, J. & Cunnings, I. (2013). The time course of morphological processing in a second language. *Second Language Research, 29*(1), 7-31.

Conklin, K. & Pellicer-Sánchez, A. (2016). Using eye-tracking in applied linguistics and second language research. *Second Language Research, 32*(3), 453-467.

Cunnings, I., Patterson, C. & Felser, C. (2014). Variable binding and coreference in sentence comprehension: Evidence from eye movements. *Journal of Memory and Language, 71*, 39-56.

Cunnings, I. & Sturt, P. (2014). Coargumenthood and the processing of reflexives. *Journal of Memory and Language, 75*, 117-139.

Dank, M., Deutsch, A. & Bock, K. (2016). Resolving Conflicts in Natural and Grammatical Gender Agreement: Evidence from Eye Movements. *Journal of Psycholinguistics Research, 44*, 435-467.

Davies, Mark. (2008-) *The Corpus of Contemporary American English (COCA): 600 million words, 1990-present*. Available online at https://www.english-corpora.org/coca/.

Dearborn, W. (1906). *The psychology of reading.* New York: Science Press.

Doherty, A. & Conklin, K. (2017). How gender-expectancy affects the processing of "them." *The Quarterly Journal of Experimental Psychology, 70*(4), 718-735.

Duchowski, A. (2007). *Eye Tracking Methodology: Theory and Practice* (2nd edition). Springer.

Dussias, P. (2010). Uses of Eye-Tracking Data in Second Language Sentence Processing Research. *Annual Review of Applied Linguistics, 30*, 149-166.

EyeLink Data Viewer 4.1.1 [Computer software]. (2019). Mississauga, Ontario, Canada: SR Research Ltd.

Felser, C. & Cunnings, I. (2012). Processing reflexives in a second language: The timing of structural and discourse-level constraints. *Applied Psycholinguistics, 33*, 571-603.

Felser, C., Sato, M., & Bertenshaw, N. (2009). The on-line application of binding Principle A in English as a second language. *Bilingualism: Language and Cognition, 12*(4), 485-502.

Fender, M. (2003). English word recognition and word integration skills of native Arabic- and Japanese-speaking learners of English as a second language. *Applied Psycholinguistics, 24*, 289-315.

The Institute of International Education (IIE). 2019. *Open Doors Report: 2019 Fast Facts.* Retrieved from: https://www.iie.org/Research-and-Insights/Open-Doors/Fact-Sheets-and-Infographics/Fast-Facts.

The Institute of International Education (IIE). 2019. *Open Doors 2019 Fact Sheet: Alabama.* Retrieved from: https://www.iie.org/Research-and-Insights/Open-Doors/Fact-Sheets-and-Infographics/Data-by-State-Fact-Sheets.

Jordan, T. et al. (2014). Reading direction and the central perceptual span: Evidence from Arabic and English. *Psychon Bull Rev, 21*, 505-511.

Just, M. & Carpenter, P. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review,* 87(4), 329-254.

Kennison, S. & Trofe, J. (2003). Comprehending Pronouns: A Role for Word-Specific Gender Stereotype Information. *Journal of Psycholinguistic Research, 32*(3), 355-378.

Kreiner, H., Sturt, P. & Garrod, S. (2008). Processing definitional and stereotypical gender in reference resolution: Evidence from eye-movements. *Journal of Memory and Language, 58*, 239-261.

Martin, K. (2011). *Reading in English: a Comparison of Native Arabic and Native English Speakers* (Master's thesis). University of Pittsburgh, Pittsburgh, PA.

Molinaro, N., Su, J. & Carreiras, M. Stereotypes override grammar: Social knowledge in sentence comprehension. *Brain & Language 155-156*, 36-43.

Nicol, J. & Swinney, D. (2003). The psycholinguistics of anaphora. In A. Barss (ed.), *Anaphora: A reference guide*, pp 72-104. Oxford: Blackwell.

Paterson, K., Almabruk, A., McGowan, V., White, S. & Jordan, T. (2015). Effects of word length on eye movement control: The evidence from Arabic. *Psychon Bull Rev, 22*, 1443-1450.

Reali, C., Esaulova, Y., & Von Stockhausen, L. (2015). Isolating stereotypical gender in a grammatical gender language: Evidence from eye movements. *Applied Psycholinguistics, 36*, 977-1006.

Roberts, L., Gullberg, M. & Indefrey, P. (2008). ONLINE PRONOUN RESOLUTION IN L2 DISCOURSE: L1 Influence and General Learner Effects. *Studies in Second Language Acquisition, 30,* 333-357.

Roberts, L. & Siyanova-Chanturia, A. (2013). USING EYE-TRACKING TO INVESTIGATE TOPICS IN L2 ACQUISITION AND L2 PROCESSING. *Studies in Second Language Acquisition, 35*, 213-215.

Sagarra, N. & Ellis, N. (2013). FROM SEEING ADVERBS TO SEEING VERBAL MORPHOLOGY: Language Experience and Adult Acquisition of L2 Tense. *Studies in Second Language Acquisition, 35*, 261-290.

Siyanova-Chanturia, A. (2013). Eye-tracking and ERPs in multi-word expression research: A state-of-the-art review of the method and findings. *The Mental Lexicon, 8*(2), 245–268.

Spinner, P., Gass, S. & Behney, J. (2013). ECOLOGICAL VALIDITY IN EYE-TRACKING: An Empirical Study. *Studies in Second Language Acquisition, 35*(2), 389-415.

SR Research Experiment Builder 1.10.165 [Computer software]. (2011). Mississauga, Ontario, Canada: SR Research Ltd.

Sturt, P. (2003). The time-course of the application of binding constraints in reference resolution. *Journal of Memory and Language, 48*, 542-562.

Thompson-Panos, K., & Thomas-Ružić, M. (1983). The least you should know about Arabic: Implications for the ESL writing instructor. *TESOL Quarterly, 17,* 609-623.

Tuninetti, A., Warren, T. & Tokowicz, N. (2015). Cue strength in second-language processing: An eye-tracking study. *THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY, 68*(3), 568-584.

The University of Alabama. (2019). International Student & Scholar Report 2019. Retrieved from the University of Alabama Capstone International Center International Student & Scholar Services website: http://international.ua.edu/facts-figures/.

Van Assche, E., Duyck, W. & Brysbaert, M. (2013). VERB PROCESSING BY BILINGUALS IN SENTENCE CONTEXTS: The Effect of Cognate Status and Verb Tense. *Studies in Second Language Acquisition, 35*, 237-259.

Winke, P., Godfroid, A. & Gass, S. (2013). INTRODUCTION TO THE SPECIAL ISSUE: Eye-Movement Recordings in Second Language Research. *Studies in Second Language Acquisition, 35*, 205-212.

APPENDIX A

Participant no. _____

What is your age?

What is your gender? Please circle one.

Male Female Nonbinary Prefer not to say

What is your current academic major?

In which country or countries have you lived? If more than one, please list them from earliest to latest.

What is your first (native or home) language?

If your first language is not English, when did you begin to learn English? Please choose one.

A. Under 6 months ago

B. Between 6 months and 1 year ago

C. Between 1 and 2 years ago

D. Between 2 and 5 years ago

E. More than 5 years ago

APPENDIX B

1. Yesterday the politician/king left the city after reminding himself/herself about the letter.
2. Normally the politician/king sat quietly while preparing himself/herself to give the speech.
3. This morning the librarian/grandmother quit the job after reminding himself/herself about the argument.
4. Even the librarian/grandmother could offer good ideas after reminding himself/herself about the details.
5. Eventually the coach/gentleman found the group after teaching himself/herself the directions.
6. Clearly the coach/gentleman made many tests before convincing himself/herself about these results.
7. Initially the assistant/girlfriend pleaded guilty without arming himself/herself with the facts.
8. Finally the assistant/girlfriend prepared dinner after teaching himself/herself the new recipes.
9. A month ago the farmer/male moved to a new country after convincing himself/herself about the relationship.
10. Yesterday the farmer/male dressed neatly before driving himself/herself to the party.
11. Luckily the nurse/lady brought cold water immediately after reminding himself/herself about the complaints.
12. Today the nurse/lady turned off the computer before introducing himself/herself to the new worker.
13. Fortunately the general/grandfather waited outside the house after upsetting himself/herself about the fire.
14. This evening the general/grandfather thought carefully while teaching himself/herself about the complicated machine.
15. Yesterday the mom/cleaner revealed everything after convincing himself/herself about the evidence.
16. In church the cleaner/mom organized flowers while teaching himself/herself to make the designs.
17. Recently the uncle/hunter worked alone after reminding himself/herself about the competition.
18. Typically the hunter/uncle operated carefully after reminding himself/herself about the strict rules.
19. Naturally the secretary/female fell in love while reminding himself/herself of the romantic letters.
20. Often the secretary/female spoke honestly while expressing himself/herself at the meeting.
21. This morning the soldier/husband left the country after blaming himself/herself for the fight.
22. Apparently the soldier/husband wrote his ideas down before involving himself/herself in the discussion.
23. Last week the victim/sister left early while reminding himself/herself about the appointment.
24. Last night the victim/sister read the book happily after preparing himself/herself a cup of coffee.
25. Obviously the killer/spokesman demanded more money after arming himself/herself with the formal letter.
26. This time the killer/spokesman entered the house after convincing himself/herself that the gun was fake.
27. At last the shopper/queen apologized kindly after persuading himself/herself about the claims.
28. Finally the shopper/queen solved the problem after encouraging himself/herself to think carefully.
29. Apparently the guard/chairman negotiated carefully while putting himself/herself at risk.
30. Today the guard/chairman investigated everybody while reminding himself/herself about the questions.
31. Today the actress/dancer practiced carefully after hurting himself/herself during the show.
32. Of course the actress/dancer expected success after reassuring himself/herself about the songs.

APPENDIX C

Please circle or cross out every word you remember seeing during the experiment.

Writer	Singer	Shopper	Celebrity	Dentist
Jurist	Sculptor	Nun	Assistant	Salesperson
Politician	Cheerleader	Criminal	Grandfather	Caregiver
Meteorologist	Groundskeeper	Pilot	Geologist	Dancer
Female	Cashier	Florist	Chairman	Athlete
Psychiatrist	Historian	Concierge	Husband	Victim
Bride	Acrobat	Toddler	Farmer	Weaver
Nutritionist	Philanthropist	Hunter	Actress	Lady
Uncle	Teenager	Cosmetologist	Doctor	Librarian
Judge	Screenwriter	Drummer	Magician	Fellow
Boxer	Cleaner	Stenographer	General	Client
Grandmother	Secretary	Student	Shoemaker	Attorney
Killer	Maid	Craftsman	Caterer	Girlfriend
Patient	Soldier	Newscaster	Mailman	Projectionist
Auctioneer	Sister	Waiter	Podiatrist	Spokesman
Supervisor	Embezzler	Landscaper	Parent	Innkeeper
Hairdresser	Swimmer	Queen	Lawyer	Baker
Cashier	Sailor	Male	Leader	Coach
Chef	King	Gentleman	Nurse	Mom
Socialite	Artist	Gunman	Youngster	Guy
Proofreader	Guard	Sunbather	Bowler	Clown

APPENDIX D

CITI PROGRAM

Completion Date 17-Apr-2019
Expiration Date 16-Apr-2022
Record ID 28331793

This is to certify that:

Elena DeCook

Has completed the following CITI Program course:

Not valid for renewal or certification through CME. Do not use for TransCelerate mutual recognition (see Completion Report).

Human Research (Curriculum Group)
Non-Medical Investigators (Course Learner Group)
1 - Basic Course (Stage)

Under requirements set by:

University of Alabama

CITI
Collaborative Institutional Training Initiative

Verify at www.citiprogram.org/verify/?wcde0d2eb-ab70-47ee-9634-50fa82266bb9-28331793

APPENDIX E

THE UNIVERSITY OF ALABAMA | Office of the Vice President for Research & Economic Development
Office for Research Compliance

October 1, 2019

Elena DeCook
Department of English
College of Arts and Sciences
Box 870244

Re: IRB # 19-OR-216, "Processing Grammatical and Stereotypical Gender Conflicts in Arabic-Speaking Learners of English: Evidence from Eye Movements"

Dear Ms. DeCook:

The University of Alabama Institutional Review Board has granted approval for your proposed research. Your application has been given expedited approval according to 45 CFR part 46. Approval has been given under expedited review category 4 as outlined below:

(4) Collection of data through noninvasive procedures (not involving general anesthesia or sedation) routinely employed in clinical practice, excluding procedures involving x-rays or microwaves.

The approval for your application will lapse on September 30, 2020. If your research will continue beyond this date, please submit the continuing review to the IRB as required by University policy before the lapse. Please note, any modifications made in research design, methodology, or procedures must be submitted to and approved by the IRB before implementation. Please submit a final report form when the study is complete.

Please use reproductions of the IRB approved consent form and recruitment flyer.

Good luck with your research.

Sincerely,