6-15-2021

An Eye-Movement Examination of Gender Stereotype Processing

Elva Adriana Garcia

AN EYE-MOVEMENT EXAMINATION OF GENDER STEREOTYPE PROCESSING

A Thesis

by

ELVA ADRIANA GARCIA

Submitted to Texas A&M International University
in partial fulfillment of the requirements
for the degree of

MASTER OF SCIENCE

December 2018

Major Subject: Psychology

AN EYE-MOVEMENT EXAMINATION OF GENDER STEREOTYPE PROCESSING

A Thesis

by

ELVA ADRIANA GARCIA

Submitted to Texas A&M International University
in partial fulfillment of the requirements
for the degree of

MASTER OF SCIENCE

Approved as to style and content by:

Chair of Committee,	Dr. Roberto R. Heredia
Committee Members,	Dr. Anna B. Cieślicka
	Dr. Monica E. Muñoz
	Dr. Daniel J. Mott
Head of Department,	Dr. Jose Carlos Lozano

December 2018

Major Subject: Psychology

ABSTRACT

An Eye-Movement Examination of Gender Stereotype Processing (December 2018)

Elva Adriana Garcia, Bachelor of Arts in Psychology, Texas A&M International University;

Chair of Committee: Dr. Roberto R. Heredia

The purpose of the present study investigates how bilinguals comprehend gender, especially when one of their languages is grammatically marked in gender. This study explored influences of world knowledge, lexical representations, and language activation in a bilingual population. Eye-movements were recorded as bilinguals read sentences containing gender-stereotyped nouns (e.g., *surgeon*) followed by a male or female pronoun (*he/she*) that served as an anaphoric referent. Language dominance in Spanish and English dominant and balanced bilinguals, the gender of the stereotype noun (i.e., female or male), and congruency between noun and anaphor (match or mismatch) were included as variables. The following eye movements were measured: *gaze duration*, *regressions*, and *right-bounded durations*. *Gaze duration* showed slower reading times for Spanish dominant bilinguals. *Regressions* to the antecedent also showed a greater effect for Spanish dominant bilinguals. *Right-bounded duration* showed longer reading times in the anaphor region and also was found higher for the Spanish dominant bilinguals. The results are supported by an inferential view that states readers infer gender through a mental representation of previously stored gender information. The associations made between Spanish, a language that specifies gender explicitly (e.g., *la mesa*; Heredia, Cieślicka, & Altarriba, 2016), and English, a language that does not contain a grammatical gender in nouns. The conditions manipulated the

pronoun to match or mismatch its stereotyped antecedent. Reading difficulties were quantified as mismatch effect by tracking eye-movements as the participants read each sentence. The anaphor and antecedent regions of interest (ROI) were measured for early- and late-stage effects through eye measurement recordings. Analyses were used to separate the effects in each of the two ROIs. Effects are reported for the early-stage processing through gaze duration and skip percentages, and for late-stage processes through right-bounded duration, first-pass, and regression times. The role of language dominance of bilingual participants in relation to the comprehension of gender in discourse is reported. Bilinguals who were Spanish-dominant showed more processing difficulty effects than English-dominant and balanced bilinguals.

Keywords: eye-movements, language dominance, bilingual, anaphora

DEDICATION

I dedicate this thesis to my perfect daughters, who have given me more than I could ever give them. To my mom, for teaching me unconditional love and support. To my dad, who frequently fed my curiosity by answering "look it up" to my many, many questions. To my beautiful sisters, your intelligence is admirable and inspiring. To my brother, a wise voice of reason when I least expected it and most needed it. Also, to my Nala, who stayed up late and woke up extra early many times to keep me company. And to all those not mentioned here from my past, for continuing to influence and drive me to pursue good things. Thank you.

ACKNOWLEDGMENTS

I would like to thank Dr. Roberto Heredia for making this experience a memorable and learning one, way beyond the requirements! Thank you for always being so willing to share your knowledge and enthusiasm with me. I would also like to thank my committee members, Dr. Anna B. Cieślicka, Dr. Monica E. Muñoz, and Dr. Daniel J. Mott, without whom this thesis would not have been possible. Dr. Anna B. Cieślicka, you have been a true inspiration and sense of reassurance in keeping it together when things got hectic. Dr. Daniel J. Mott, thank you for always bringing your sense of humor and honesty to the lab. And Dr. Monica E. Muñoz, the beginning of my research path would not have been made a reality without you. Also, I'd also like to express my gratitude to Mr. Marcus Johnson for all his generosity in the eye-tracking data analysis. The journey would not have been possible without a single one of you, and I consider myself fortunate to know a group of brains such as yourselves. Thank you.

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

INTRODUCTION

As a way to introduce the topic, consider a highly opinionated individual exhibiting confidence and assertiveness. How would this individual be evaluated? How might an assertive individual exhibiting confidence be evaluated? One possibility, of course, is that *assertiveness* might be *in the eyes of the beholder* since a female exhibiting such trait might be described differently than a male. Case in point, the 2018 fallout at the US Open Tennis Tournament of Serena Williams' "assertiveness" challenging the umpire and being portrayed as "angry." Although it is not clear if the approach taken to assess Serena Williams' attitude was due to her ethnicity, gender, or both, other cases in the same domain point to a possibility that gender stereotypes cannot be ruled out. Briefly, stereotypes are unconscious representations that operate automatically to make generalized categorizations and judgments about others (Devine, 1989; Greenwald & Banaji, 1995; Katz & Braly, 1935, p. 181). Stereotypes are specific beliefs attributed to group members of a particular social group (e.g., gender, ethnicity), and these stereotypical representations are held at implicit levels that are readily accessible under appropriate contextual conditions (Carreiras et al., 1996; McKoon & Radcliff, 1992). Likewise, gender stereotypes are specific to how men and women are categorized. For instance, women are viewed and expected to be compassionate, while men are associated with competitiveness and assertiveness (e.g., Fiske, Cuddy, & Glick, 2007; Heilman & Eagly, 2008). Various occupations (e.g., *surgeon*/*nurse*) have also been characterized as mostly female or male (Kennison & Trofe, 2003). How do we comprehend stereotyped language? More specifically, how do we process gender-stereotyped language information during the reading comprehension process? The present study investigates

---

This thesis follows the model of *Language & Cognitive Processes*.

gender-stereotype processing by measuring participants' eye-movements as they read sentences with biased male-to-female contextual information. One purpose of the present study is to investigate gender-stereotyped processing using a population whose first (L1) or second (L2) language is grammatically marked in gender. A second purpose of this thesis was to compare differences, if any, between language dominance in English and Spanish and its influence on gender in occupational role nouns during the reading comprehension process. We employ an eye-movement paradigm to address issues in gender agreement and the different linguistic levels of representation during reading comprehension. We begin this discussion by providing an overall view of Spanish and English in relation to conceptual and grammatical gender marking. We then describe current theoretical formulations and general research findings from the literature on social stereotypes. We now review some general gender-related linguistic aspects of Spanish and English.

Spanish, unlike English, marks objects in terms of gender. The noun *table,* for example, is classified as gender neutral in English but would be a feminine-gendered noun in Spanish (e.g., *la mesa*). Even when concepts in English are not explicitly marked in gender (e.g., *doctor*, *barrel*), the Spanish rule would grammatically mark such cases as having a feminine or masculine gender (e.g., *la doctora*, *el doctor*, *el barril*). Shifts in gender roles, occupational roles, and societal attitudes have taken a more inclusive and progressive approach. This has spilled over into areas such as the use of languages, especially those with grammatical gender like Spanish. This movement's effort is to do away with gendering nouns, in particular, occupational roles (e.g., *policeman* to *officer*). A noun in English that indicates its gender is not only explicit but is represented in the mental lexicon. In nouns such as *policeman,* gender-specificity (i.e., *man*) is part

of the meaning and are processed lexically. Accordingly, *man* must be a male by definition, but, together, *policeman* also automatically categorizes all officers as male, at least before the possibility of a female officer is considered. If advocating for a preference of un-gendering languages, as in replacing *he*/*she* for *we*/*they*, would an underlying connotation make an observable bias during the perception of these types of nouns? Is using a gender-free noun such as *police officer* prevent the reader or listener from concluding that the office must be a male based on one's personal experience or social influences? As for Spanish and other languages that mark gender grammatically, occupational role nouns, such as *la doctora* (female doctor) or *la abogada* (female lawyer), clearly specify the biological sex of the noun, female in this case. (Note that some animals in Spanish such as *zopilote/buitre* [*vulture*], and *mapache* [*raccoon*] connote maleness; The same is true for Polish). Therefore, Spanish terms grammatically labeled as feminine (e.g., *La Latina*) or masculine (e.g., *El Latino*) are adhering to the shift by replacing the specifying morpheme that grammatically marks and denotes the expectation (e.g., *a*/*o*) with a gender-neutral one (e.g., *x*; *Latinx,* or *Chicanx*). This sociolinguistic phenomenon has been currently adopted by major U.S institutions of higher education such as Amherst College (Latinx & Latina American Studies), University of Oklahoma (Latinx Studies), and University of Missouri-Kansas City (Latinx and Latin American Studies Program), for example.

Further, if bilinguals possess the conceptual and grammatical representation of gender-based classification, how is gender-typical information that does not specify gender in English (e.g., *surgeon* or *nurse*) processed? Are bilinguals more likely to exhibit a larger mismatch cost (i.e., take longer times to comprehend) as they encounter gender-biased occupational roles that might conflict with their schematic representation that, for example, a *surgeon* in English must be

a male because *cirujano* in Spanish is more likely to be more frequently encountered (based on word counts on written material) than *cirujana*? If so, this would have implications for the role of language dominance and fluency in L2 during comprehension in gender discourse. That is, if Spanish-dominant bilinguals show greater mismatch costs than English-dominant or balanced bilinguals, the implication would be that Spanish, a language that explicitly reveals the gender of the noun, influences the reader by providing access to the Spanish representations of the concept under consideration.

GENDER INFORMATION AND LEVELS OF REPRESENTATION

What is the role gender information plays in the reading comprehension process as it relates to grammatical/conceptual agreement between the subject noun and pronoun at the sentence level? Gender-based information is represented at different linguistic levels (i.e., lexical, morphological, and conceptual). Nouns explicitly marked in gender assign a male or female role and abide by constraints of gender agreement. For instance, the concept of *mother* is represented in gender at the lexical level, and its referent is specified to be female; that is, *she* would be the agreeing pronoun that follows. Too, *princess* or *actress*, are marked in gender by their morphological structure (e.g., "ss") denoting a female. However, other nouns like *surgeon* are conceptual in gender, meaning they do not explicitly assign a gender. A surgeon can be either a female or male; however, from a worldview perspective, the construction of the mental representation of this occupation would be typically masculine. Through world knowledge, gender is processed as a construct via certain schematic representations. The social perspective of gender differs from biological sex (i.e., male vs. female). Sex is anatomically defining (e.g., reproductive organs) while gender is a social construct defined by the perspective view of society and expressed by social

assignment. Gender is dichotomously represented, then, in marked characteristics of female and male in social roles (West & Zimmerman, 1987), including occupations.

Current theoretical views posit that gender representation is a social adaptation and part of what is learned and stored in one's mental lexicon. One view, for example, proposes that gender information is accessed through real-world knowledge and biases the reader's commitment to infer gender accordingly. As readers attempt to make sense of nouns with unspecified gender, they must rely on their knowledge to make inferences about the corresponding gender (Carreiras et al., 1996; Kennison & Trofe, 2003; Oakhill, Garnham, & Reynolds, 2005). Other views, in contrast, suggest the absence of gender agreement may create a lexical interruption that can be overridden by context when it does not match the appropriate conditions. Current research addresses how readers process gender under contextual conditions that create gender ambiguities, such as lexical and conceptual nouns, and the time taking to resolve the prospective lexical ambiguity. That is, when does gender information become accessible and applied during reading comprehension? The *how* and *when* questions are of primary interest because the answers would not only offer a more precise location (i.e., early vs. late accessibility) of stereotypical information roles, but also establish the importance of stereotype construction during the thought process (Sanford & Garrod, 1989). Further, differences in whether stereotypical gender is processed in the same way as lexical gender would have implications for differences in the levels of mental representations.

ANAPHORIC PRONOUN RESOLUTION

Nouns and pronouns, in English, are part of a natural language system where gender agreement is expressed between nouns and pronouns without depending on other modifiers, such as definite articles or determiners, as they do in languages marking with grammatical gender (e.g.,

the, a vs. Spanish *la/el, una/uno*). English pronouns match the gender of a referent subject noun (e.g., *girl ~ she*) in discourse to make sense grammatically. Evidence from definitional gender noun (e.g., *king*/*queen*) experiments have revealed that these nouns are unmovable in terms of how they are processed and interpreted. That is, they are not easily modulated by context order (i.e., the location of the pronoun), morphological or grammatical dependencies, other than their syntactic constraints that explicitly assign gender.

Unlike lexically and morphologically represented nouns, conceptually represented nouns are not explicitly marked and are not held to grammatical constraints. One way in which conceptual nouns are assigned gender is through anaphora (Forbes, Poulin-Dubois, Rivero, & Sera, 2008; Greene, McKoon, & Ratcliff, 1992). Briefly, anaphora or anaphoric resolution refers to the process in which a referential pronoun (e.g., *he*) follows a referenced subject or antecedent in the sentence (Rayner, Pollatsek, Ashby, & Clifton, 2012). Thus, for the sentence, *The professor is fun but he is difficult*, "professor" would be the antecedent, and the pronoun *he* would be the anaphor. During the reading comprehension process, the reader must make the inference that the pronoun *he* must referred back to the *professor*.

Empirical evidence shows a consistent pattern on how gender-based information influences the reader's comprehension beyond the typical natural gender agreement that would come with processing dependencies (Carreiras et al., 1996; Duffy & Keir, 2004; Garrod & Sanford, 1982; 1994; Hillert & Nakano, 2016; Kennison & Trofe, 2003; Kreiner, Sturt, & Garrod, 2008). For the sentence, *The girl was at school because she was a student*, the natural gender expectation specifies that *the girl* (the noun phrase [NP]) and *she* (the pronoun) must agree in gender to be clearly understood. Whereas, if the sentence, *The doctor was out because she was sick* followed the same

natural gender principle, *doctor* would draw an equal or neutral expectation of *he* or *she* because the referent could be either male or a female. Yet, sentences containing nouns with an unspecified gender (e.g., *doctor*/*nurse*) are processed and comprehended differently in comparison to nouns with explicit gender (e.g., *girl, princess*). In the case of *girl*, the level of representation is a lexical one because female is part of its definition. However, a conceptual noun, such as *doctor*, would be represented at an inferential level since no grammatical gender cue is present. The noun *doctor* would be considered a male, stereotypically, which means that the probability for a male referent would be higher than a female one. Thus, this creates an expectation for the pronoun *he* when referring to the conceptual noun *doctor* in a sentence, even if it is momentary.

The role of gender agreement in processing dependencies influences the reader in drawing inferences from contextual information. To comprehend sentences and its components, gender agreement of the NP constrains upcoming words, such as the referential pronoun (e.g., Carreiras et al., 1996; Kennison & Trofe, 2003; Kreiner et al., 2008; Oakhill et al., 2005; Van Berkum, Brown, & Hagroot, 1999). As an example, consider *king*, a noun lexically defined in gender that is presented as the subject referent in a sentence. The referential pronoun following the noun would be *he* because both are male, agreeing in gender. Although gender is explicitly specified in definitional nouns, it is in conceptual nouns (e.g., *surgeon*) that the agreement can become arbitrary (Kreiner et al., 2008; Osterhout, McLaughlin, & Bersick, 1997). That is, when a conceptual noun is encountered in context, the expectation of the reader can be biased as he/she decides on a gender for the subject noun, and along with it, its anaphoric pronoun. In language processing, conceptual nouns are argued to be stereotypical and inferred using world knowledge as a primary basis.

To illustrate, consider Sanford's (1985) discourse riddle: *A man and his son were away for a trip. They were driving along the highway when they had a terrible accident. The man was killed outright but the son was alive, although badly injured. The son was rushed to the hospital and was to have an emergency operation. On entering the operating theatre, the surgeon looked at the boy, and said, "I can't do this operation. This boy is my son." How can this be?* (p. 311). How can that be possible if the father died in the accident? Indeed, this demonstrates that even when discourse is constrained by the father's death, a reader more likely experience processing difficulties to override the typicality of a male surgeon (Garnham, 2001, p. 142). Some studies have suggested that biases held interpretations weigh on the reader to commit toward the gender that most matches that bias (Van Berkum et al., 1999; Garnham, 2001; Reynolds, Garnham, & Oakhill, 2006; Kreiner et al., 2008; Gygax, Gabriel, Sarrasin, Oakhill, & Garnham, 2009). Other explanations have attempted to pinpoint the representation of a gender-ambiguous noun and whether it is represented at the lexical, or inferential level (Cain & Oakhill, 1999; Garnham, 1981).

In a similar vein, *cataphora* has been used to further evaluate sentence structure and word order to understand how gender information is comprehended (Kazanina, Lau, Lieberman, Yoshida, & Phillips, 2007). In cataphora, the subject pronoun is presented prior to the referent noun as in the sentence, *I was told he* (subject pronoun) *was difficult, but the professor* (referent noun) *turned out to be very nice*. Pronoun resolution time-course studies (see below) have revealed processing difficulties (i.e., mismatch costs) when violations in gender agreement of conceptual nouns are presented in anaphoric (e.g., Osterhout et al., 1997; Van Berkum et al., 1999; Lauro & Schwartz, 2018), but not cataphoric sentential material (e.g., Carreiras et al., 1996; Duffy & Kier, 2004; Kreiner et al., 2008). In contrast, definitional nouns produce a mismatch cost effect,

regardless of cataphora or anaphora. This evidence suggests the difference in noun types (definitional and stereotypical) may modulate the mismatch cost effects in gender agreement violations when the noun and pronoun are repositioned in the sentence. Additionally, an inferential view would support that a stereotypical noun (e.g., *nurse*/*surgeon*) in an anaphoric sentence would trigger biased expectations of the subject's gender (Kennison & Trofe, 2003; Kreiner et al., 2008). We now review theoretical models that have been used to explain how gender information is processed, in conditions in which context provides conflicting gender information during the reading comprehension.

THEORIES OF GENDER COMPREHENSION

The *mental models approach* poses that during the reading comprehension process, readers access mental representations that connect associations from prior or world knowledge (e.g., Johnson-Laird, 1983; 2010) and any available contextual information such as gender agreement cues (e.g., Garnham, 2001; Kennison & Trofe, 2003; Kreiner et al., 2008; Zwaan & Radvansky, 1998). This approach suggests that during the reading processes, incoming text information operates in a top-down processing fashion where this input is integrated through knowledge previously encoded (Bobrow & Brown, 1975; Kurby, Britt, & Magliano, 2005; McCormick, 1988) containing, to some degree, cognitive control over arriving at the interpretation regarding ambiguous text. As pointed out by Garnham (2001), mental models play critical roles during the inferential process, especially when the text content is ambiguous. Accordingly, associations are based on world knowledge information and expectations that, for example, the gender occupation *nurse* must belong to a female. In reading the occupational role of *nurse* that is stereotyped in gender, a probabilistic view of *nurse* would create an implicit bias to generalize and assign gender

for all *nurses* in future encounters (Johnson-Laird, 2010). Thus, when text content is incongruent or mismatched (e.g., *males ~ nurses*) the existing mental model (e.g., *females ~ nurses*), a reading cost would incur and be reflected in an increase of processing complexities or difficulty (as revealed by a significant increase in reading times) by the reader. For this reason, it is hypothesized that the activation of the stereotype occurs automatically at the first encounter of a stereotyped gendered-noun. This suggests that a stereotyped noun is packed with information biasing in gender recognition. The question is, *when* is it accessed? If the mismatch cost is detected upon first encountering a mismatching pronoun following the gender stereotyped noun, then would it be the mental model assisting to comprehend and posing reading difficulties when a referent stereotyped noun (e.g., *surgeon*) does not match the anaphor (e.g., *she*)?

The schematic approach is a type of mental model that creates a frame of reference for any received social information (e.g., physical traits of a group). This information is retrieved and triggered upon encountering relevant stimuli (Allport, 1954; Bargh, 1994; Barlett, 1932) and is related to implicit stereotyping or thoughts assigned to and about certain social groups. Thus, social stereotypes become active in thought and behavior (Bargh, 1994). They involve automatic thought processes either at the point of activation or observed as a particular experience/concept that does not fit the schematic representation of the stereotype. As argued by Zwaan and Radvansky (1998), a schema enables comprehension from information representations and the situational model constructed by the reader (Van Dijk & Kintsch, 1983). This situational model is a mental picture of what is on the text which then draws from world knowledge information to comprehend context with nouns that contain gender stereotypical information (Carreiras et al., 1996).

Moreover, from a social evolutionary perspective, the *us vs. them* view in relevant social settings is highly attractive (Tajfel & Turner, 1979). Social evolutionary theories focus on survival as a goal. The driving force behind this goal is the outcome that guides implicit behaviors, such as being guarded of surroundings and maintaining a certain mindset toward anything that is different. Although this view has been applied mostly to racial/ethnic themes, it applies to studies of gender stereotyping as well (e.g., Deaux, 1995). This view posits that humans are conditioned to think categorically about others in terms of *in*- vs. o*ut-group*. Accordingly, males view males as equals (*in-group*) and females as unequal (*out-group*) and vice-versa. In one experiment, Banaji and Hardin (1996) presented male and female participants with stereotypical role nouns (prime) paired to a personal pronoun (target) in their reflexive (e.g., *himself*/*herself*) or a possessive form (i.e., *his*/*hers*, *he*/*she*). They measured the automaticity of stereotypes in in a reaction time study by manipulating the pronoun presentation with the gender typicality of the noun. In the *prime-target* pair *mechanic-she,* the condition would be a mismatch. The experiment presented word pairs and participants decided whether the target (pronoun) was a male or female pronoun. Their results found support for the activation of gender information. Participant responses were faster when gender stereotyped role nouns (e.g., *mechanic*) matched with a male pronoun (e.g., *him*/*he*/*his*). However, Banaji and Hardin's results are not clear, since it is difficult to determine if the activation of the stereotype occurred during prime or target onset. That is, their findings were limited to stereotype activation but did not clearly determine whether they were accessed lexically or inferentially. It is also possible that the task used was not as implicit as they claimed. In this case, the task might have encouraged strategic processing.

The *inferential view* poses that inferences are made through probabilistic algorithms based on world knowledge. That is, gender assignment is produced by the probability and expectation of a gender-biased assumption about the stereotyped noun. In resolving gender ambiguity, this view suggests that gender ambiguous nouns are assigned gender when they precede an anaphoric pronoun or reflexive in discourse or sentence context (Kreiner et al., 2008). In an anaphoric sentence, a gender-stereotyped noun is unspecified in gender but is followed by a gender-congruent anaphor. For instance, the referent *nanny* is presumed to be female because of the stereotype; therefore, the expected anaphoric reference would also be female or *she*; however, when it is followed by a mismatching pronoun (e.g., *he),* it would produce a conflict between thought and behavior (e.g., regressions in eye-movement measurements) and delays in reading processing times.

As argued by Rayner et al. (2012), drawing inferences from text, as when the meaning is not explicitly stated, is necessary to make coherent interpretations. As an example, consider the following sentences (p. 265):

Read sentences:

(1) *Keith took his car to London (Explicit).*

(2) *Keith drove to London (Implicit).*

Test Sentence:

The car kept overheating.

This demonstration examines whether the *car* that kept overheating would be inferred as pertaining to the car that was driven by Keith (i.e., *his*) in sentence (1). A *car* was mentioned only in the first sentence. However, results showed that readers processed both sentences at the same

speed. These results were interpreted as suggesting that gender information did not require the mental lexicon to access information about the stereotype; accordingly, gender-stereotyped information required higher levels of processing (e.g., problem-solving) to infer the referent's sex since these nouns were not explicitly marked with grammatical gender cues (Oakhill et al., 2005; Kreiner et al., 2008). Other studies involving anaphora and cataphora in gender comprehension have also revealed that additional processes are required to make inferences through knowledge outside of semantics.

Lexical access theories propose that a reader accesses words through the mental lexicon and is compelled to conform to syntactic constraints as in gender agreement. This approach to reading comprehension assumes that a bottom-up processing system drives the interaction between text information coming in to find a possible and probable meaning (Mason & Just, 2007; Marslen-Wilson & Welsh, 1978; Swinney, 1979). It assumes a system in which incoming information is managed and directed by rigid signals that are learned (or established) and maintained to be fitting as if following a template on each word (Waltz, 1975).

Accordingly, gender is accessed lexically, which means that a word is processed and comprehended at each point it is encountered, regardless of its position in a sentence (Rappaport, Levin, & Laughren, 1993). That is, if a referent's gender (*he/she*) is alluded to in earlier context, a subsequent subject noun would be expected to be of one gender more than the other. This expectation has been observed in reading studies, such as reaction time costs (Carreiras et al., 1996), eye-tracking (Kreiner et al., 2008), and event-related potentials (ERPs; Osterhout et al., 1997). Briefly, ERPs are measured brain responses by electroencephalography (EEG) that tell of an event through waveforms (e.g., P600) with one of the highest temporal resolution in

experimental measurements. If discourse presents a reflexive (*himself*/*herself*; Duffy & Keir, 2004) or a pronoun (e.g., Kreiner et al., 2008) before mentioning an ambiguous occupational noun (e.g., *minister*), readers would process a mismatch much faster in this order of presentation (i.e., cataphoric). Encountering the stereotypical role noun in prior discourse would create an opportunity for the reader to experience difficulties if the pronoun does not match its stereotypical gender. If the stereotyped noun is presented anaphorically, a mismatch cost would be incurred, revealing that conflicting gender information would bias the expectation of a subject's gender. Carreiras et al. (1996), Osterhout et al. (1997), and Kreiner et al. (2008) looked at lexical activation using definitional nouns that included gender in their definition (e.g., *girl*), nouns based on stereotypes (e.g., *minister*), and nouns based on social roles. These researchers compared grammatical vs. stereotypical role nouns using English and Spanish (Carreiras et al., 1996) and English and German (Reali, Esaulova, & Von Stockhausen, 2015). Spanish and German are languages that mark grammatical gender, while English does not.

A difference between the two studies is that Reali et al. used German descriptors stereotyped with gender cues (e.g., *-in*; *Tischler*, masculine/*tischlerin*, feminine) instead of role nouns that might otherwise explicitly state the occupation of the subject. In contrast, Carreiras et al. looked at stereotypical nouns in English (e.g., *nurse*) and in Spanish (e.g., *la enfermera*) to test whether the role of a definite noun with morphological gender cues in Spanish (i.e., *a/o*, *la*/*el*) made a difference. Carreiras et al.'s results indicated that the stereotypical nouns in English elicited reading difficulties in the mismatching conditions. For the Spanish mismatching conditions, they found that readers would only show the mismatch cost effect at the first encounter of a stereotypical mismatch (e.g., *la carpintera*) but would show no other difficulty in the second anaphoric reference

(e.g., *ella*). This finding was interpreted as being due to processing at the definite article's position in which gender was established prior to the stereotyped noun. The implication was that readers accessed gender-based information in earlier stages of language processing which they measured by reading time differences.

Osterhout et al. (1997) compared and contrasted the elicited event-related potential (ERP) waves elicited as readers encountered sentences presenting definitional and stereotypical type nouns. They expected a difference in brainwave activity between the two noun types since each noun was thought to be represented at different linguistically levels. Definitional nouns were hypothesized to elicit a P600 ERP while the stereotypical nouns were supposed to elicit an N400 ERP. A P600 effect is a positive wave considered to be a syntactic effect that is present with anomalies in syntax, such as grammatical errors (e.g., *mother ~ him*). The N400 is a negative wave associated mainly with pragmatic responses, and not with grammatical structure violations (Kutas & Federmeier, 2000) such as violations in gender agreement of definitional nouns. In Osterhout et al.'s (1997) experiment, readers resolved the stereotypical noun's gender with the anaphoric reflexive (*himself*/*herself*) that was either matched (*nurse ~ herself*) or mismatched (*nurse ~ himself*) to the stereotype. For the definitional nouns, a similar sentence (1) structure was presented also matching (*father ~ himself*) and mismatching (*father ~ herself*) in gender agreement.

(1) *The man prepared himself/herself for the interview*.

(2) *The doctor prepared himself/herself for the operation*.

Since stereotypical nouns are conceptual, theoretically speaking, no grammatical constraints exist in the comprehension process. For example, no morphological markings (e.g., *princess* vs. *prince*) would dictate the gender of conceptual nouns such as *doctor*. Therefore,

Osterhout et al. considered the possibility that an N400 would be elicited in the mismatching stereotypical noun conditions (2) instead of a P600. However, a P600 was recorded as readers encountered mismatching gender information in the definitional noun conditions but were also recorded in the stereotypical conditions. To summarize, their results supported the lexical accessibility of stereotypical information as being processed through a similar level of linguistic representation as definitional or grammatical gender. That is when gender agreement violations are encountered in reading definitional (e.g., *father ~ herself*) or stereotypical (e.g., *minister ~ herself*) the effect of processing difficulty is similar as reflected by the P600 found by Osterhout et al.

In another study addressing the lexical view, Duffy and Keir (2004) used discourse with context that resolved the ambiguity prior to stereotype presentation with sentential information that conflicted world knowledge information (e.g., *babysitter ~ himself*). They presented anaphoric sentences with stereotypical nouns and a reflexive pronoun as in, *The babysitter found himself humming while walking up to the door*. Their results suggested that a stereotyped noun, like a definitional noun, was activated immediately and automatically at any, and every, position within the discourse, whether or not prior context had established the referent's gender (see for example, Jurafsky, 1996). Their results also supported the idea that conceptual nouns that are stereotypically female or male have properties that are context-independent (Duffy & Keir, 2004; Kintsch, 1988), meaning that the noun acts independently, or alone, as it is being interpreted. Thus, if the noun *surgeon* is presented in context or by itself, the access of its meaning and any integration required referring to that noun would be constant and represented as male (according to stereotypical information) each time. It follows, then, that a stereotypical noun would be considered to fully

access information in a similar way as a definitional noun accesses the mental lexicon (i.e., *king*); so that when integrating the anaphoric pronoun, the only way to not violate the gender information would be to include the pronoun *he* and never *she*.

As argued, mismatch costs incur when the pronoun does not match the stereotype in sentence comprehension; however, as per the lexical view, if the context establishes gender early on in the discourse, the activation of the stereotype is decreased (or suppressed) resulting in less to no interference (Duffy & Keir, 2004; Hess, Foss, & Carroll, 1995) and faster processing (Carreiras et al., 1996; Kreiner et al., 2008). More specifically, when experiments introduce and compare a cataphoric vs. anaphoric presentation of the stereotype (Carreiras et al., 1996; Kreiner et al., 2008), the mismatch cost effect disappears for stereotypical but not for the grammatical conditions. For example, the cataphoric sentence with stereotypical information, *After reminding herself about the letter, the minister immediately went to the meeting*, would be processed with less reading difficulties than in an anaphoric sentence presentation (i.e., *The minister immediately went to the meeting, after reminding herself about the letter*) with stereotypical information, and even faster than sentences that mismatch in grammatical gender (e.g., *himself*, *mother*). The cataphoric sentence, *After reminding himself about the letter, the mother immediately went to the meeting*, contains a grammatical mismatch where the reflexive pronoun *himself* and the definitional noun *mother* do not meet the gender agreement rule at a grammatical level. When a grammatical violation (*mother ~ he*) is encountered, the mismatch cost effect is so large that there is no chance to override this cost effect regardless of whether the sentence was cataphoric or anaphoric (e.g., Carreiras et al., 1996; Kreiner et al., 2008; Osterhout et al., 1997). This overriding effect was expected for stereotypical violations (*surgeon ~ she*) for both sentence types; however,

in Duffy and Keir (2004, Experiment 2), the findings revealed that this override occurred in the second encounter of the pronoun but not in the first. Duffy and Keir's (2004) interpretation of this finding was that conceptual nouns do not contain context-independent properties in the same way that grammatically represented nouns such as *father* do, contrary to Osterhout et al.'s (1997) results. Duffy and Keir (2004) and Hess et al. (1995) further concluded that in discourse, the stereotype was activated and established at the noun's position; then, the encounter of the first presented mismatched anaphor (*he*/*she*) enabled the reader to reinterpret the stereotype to match the anaphor so that the second encounter would not elicit the mismatch cost again.

The *Lexical Reinterpretation Model* (*LRM;* Duffy & Keir, 2004) has been proposed to explain that comprehension in cataphoric discourse is maintained by updating the representation of the discourse as it is read (Hess et al., 1995). According to this view, when stereotypical gender information is encountered in a context that has established the gender of the noun, the mismatch effect is reduced, or reinterpreted. Duffy and Keir examined the predictions of LRM by employing definitional nouns to determine if, like stereotypical gender nouns, they too could be modulated in cataphora-type sentences (cf. Carreiras et al., 1996). Duffy and Keir's (Experiment 2) participants read passages such as,

Jeff's/Lucy's power had been unreliable ever since the tornado.

The electrician was a cautious woman who carefully secured her ladder to the side of the house before checking the roof.

Jeff suspected that high winds had loosened the connection to the power lines.

The electrician taught himself a lot while fixing the problem.

The passage above is an example of a mismatching condition in which the stereotypical information does not match between the social role (*electrician*) and gender (*woman*). The contextual information relates to an *electrician*, then the subject is stated as a *woman*; it is then restated by the anaphoric reflexive pronoun *her*. Duffy and Keir hypothesized that the last pronoun *her* would be read faster since the contextual information dictated that the *electrician* was a *woman* explicitly in the earlier part of the discourse. Mismatched conditions presented in discourse in this sequence would produce a reinterpretation effect, as predicted by LRM. The first presentation of the explicit gender (i.e., *woman*), regardless of the stereotyped noun (i.e., *electrician*), would modulate the gender in the subsequent context. Duffy and Keir's (2004) wanted to determine whether discourse context would affect the interpretation of context-independent nouns vs. context-dependent properties. Context-independent, in this case, refers to nouns that are integrated a certain way and remain consistent, regardless of context (Kintsch, 1988). Duffy and Keir's discourse presentation included a stereotyped noun in the first line, followed by a reflexive pronoun, and then, for a second time, the gender pronoun in the third line. The second anaphoric pronoun was predicted to override the mismatch cost effect. Duffy and Keir proposed that the stereotype would be activated upon the first encounter of the noun (e.g., *electrician*), accessed in the first anaphor (*herself*), and overridden in the second anaphoric reference since the stereotypical mismatch had been integrated by that point in the discourse.

Results from Duffy and Keir's (2004) eye-tracking experiments showed immediate activation of gender information at the first mention of stereotyped nouns (i.e., referent) and the processing difficulties decreased in the second encounter for the mismatching conditions of the stereotypical information. This resulting effect was interpreted as a byproduct of the presentation

of the subject noun's gender that established gender previously. Duffy and Keir's results illustrate how a stereotyped noun is lexically accessed and does not specify an exact point of activation to ascribe a mismatch effect during reading (for mismatch conditions). However, it does loosely establish the importance of sequence presentation in discourse, as well as the capacity for stereotypical information to be overridden once it has been integrated (within discourse). In comparison to grammatical gender, which cannot be overridden even if discourse makes mention of gender disagreement, stereotypical gender has been less rigid and able to be "reinterpreted."

Overall, the theoretical models discussed above offer different perspectives on how gender information might be accessed during the reading comprehension process. Interestingly, the inferential and lexical views seem to be complementing one another instead of constituting as opposite theoretical approaches. While theories based on mental representations indicate that inferences are made automatically through accessing world knowledge primarily, others (e.g., the lexical access theory) operate under lexical constraints. Contrary to a lexical view, the level of representation is not uniformed for conceptual nouns given that they carry more world and external information that can be subtle and implicit, but close enough to the surface where it can be used to interpret gender during context comprehension.

RESEARCH FINDINGS

Carreiras et al. (1996), Kennison and Trofe (2003), Kreiner et al. (2008), and Reali et al. (2015) looked at gender agreement in sentence processing. These studies investigated the effects of cataphoric and anaphoric sentences and concluded that gender information exhibits early processing effects which in turn lead to interference as conflicting information is encountered. These findings support the role of early activation of gendered-related information as predicted by

the inferential and mental models. Here, we would like to underscore that Kreiner et al. (2008) and Kennison and Trofe (2003) used English monolingual populations, while Carreiras et al. (1996) used English monolinguals and subsequent studies utilized Spanish monolingual speakers. Likewise, Reali et al. (2015) used German-speaking participants and did not address issues related to bilingualism.

Carreiras et al.'s (1996) purpose was to explore how stereotypical gender information is comprehended in context. In addition, they looked at the theoretical issue of early- versus late-stages as it relates to the accessibility of gender-related information in English and Spanish. In Experiment 1, Carreiras et al. (1996) investigated gender processing through anaphoric discourse using English sentences to establish the mismatch effect between the stereotype (e.g., *footballer*) and its anaphor (i.e., *he*/*she*). As predicted, results showed a mismatch between the stereotypes when it conflicted with the anaphor. Further, these results indicated that the stereotype was established early in anaphoric sentences (Sentence 1: *The footballer wanted to play in the match*) by slowing down the readers as they encountered a second sentence (Sentence 2: *He/She had been training very hard during the week*) that contained the anaphor. The presentation of the pronoun referencing *footballer* in the first sentence was the manipulation to match or mismatch the stereotype of football players; thus, for the mismatch conditions, they found that readers would read the second sentence slower than in matched conditions where no stereotyped information was violated.

In Experiment 2, Carreiras et al. measured two main sentences (first and third) in Spanish to see if the first sentence (*El carpintero/La capintera tomó las medidas para hacer el armario; The carpenter took measures to build the closet*), followed by a second (*Era un cargo bastante*

urgente; It was a very urgent task) would establish the stereotype for the third sentence (*El/Ella tenía que terminarlo en el plazo de una semana; He/She had to finish it in the space of one week*). In the first sentence, *El carpintero* (male carpenter) was the stereotyped noun that was established by the preceding definite article (i.e., *el*) and morphologically by the suffix *-o* in *carpintero*. In the mismatching conditions, *La carpintera* (female carpenter), was hypothesized to produce a mismatch effect by slowing the reader down as the pronoun anaphor encountered the pronoun anaphor in the third sentence (e.g., *Ella tenía . . .; She had. . .*). In short, Spanish speakers showed the mismatch effect and no differences in reading speeds for the second sentence, since *un cargo* is not linked to the pronoun subject (*El*/*La carpintero*/*a*) once the stereotype had been processed beyond activation. These results were similar to the overriding effects reported by Hess et al. (1995) and Duffy and Keir's (2004).

Likewise, Experiment 2 showed that Spanish speakers had access to the stereotype very early during reading. It appeared that the stereotype was, in fact, activated when the stereotyped noun was first encountered in Spanish. When the gender of a definite article (*El*/*Ella*) and a morphologically marked noun mismatched its stereotypical gender, a mismatch cost (i.e., reading difficulties) incurred in the first sentence with no difference observed for the second sentence. The idea for Experiments 2 was to establish that Spanish speaking participants reading Spanish sentences had access to the stereotype earlier than English speakers in Experiment 1 reading English sentences. Accordingly, results were due to the morphological marking and the definite article that accompanied it; whereas, in the English experiment, the mismatch effect was not detected until later discourse mismatched the gender that was expected in the second sentence containing the pronoun resolution. In Experiment 3, Carreiras et al. replicated the findings from

Experiment 2 and confirmed that participants were not using strategies nor were becoming distracted by the pronouns, *el* or *ella* as they read the third sentence (e.g., *El/Ella tenía que terminarlo en el plazo de una semana; She had to finish it in the space of one week*). The first three sentences were similar to those used in Experiment 2, the third sentence contained the anaphoric reflexive pronoun, where the gender of the subject was referenced. They included a fourth text line to include the comprehension question that related to the previous discourse information (i.e., sentences 1-3) in order to establish that no other interference was occurring, for example, from the definite article in the first sentence.

In Experiment 4, Carreiras et al. further manipulated sentence presentation by separating it into two parts. For the first part of the sentence, participants pressed a computer button to read the NP (E*l carpintero; male carpenter*) while the computer measured the time it took to read and moved on to the next segment. The second part contained the continuation of the sentence . . . *tomo las medidas para hacer el armario* (*took measurements to make the cupboard*). Reading times were taken for the second part as well. The results of Experiment 4 showed an interaction between matching and mismatching conditions for stereotypical male nouns only, showing faster processing times when the noun matched the stereotypical information between definite article and noun vs. the mismatching definite article and stereotyped nouns (e.g., *la carpintera;* female carpenter) in the first part of the sentence. A similar effect was found for the second part of the sentence, even though there were no gender biased nouns or pronouns present. Reading times were also slower for the second part of the sentence that followed a stereotypical mismatch (i.e., *la carpintera;* female carpenter). Overall, what these results suggested to Carreiras et al. was that

information about the stereotype occurred very early in Spanish and somewhat similarly in English, even without the morphological marking.

Kennison and Trofe (2003) investigated stereotypical gender nouns in the context of paired sentences to examine pronoun (*he* vs. *she*) resolution by English monolinguals. Unlike Carreiras et al. (1996), Kennison and Trofe used the visual moving window (VMW) task to measure reading. Sentences were systematically parsed into segments containing regions of interest (ROIs) for words presented one at a time (Heredia, López, García, Altamira, & González, 2016). Thus, for the sentence |*The executive*| *distributed*| *an urgent memo.* | *He/She*| *made it clear*| *that work*| *would*| *continue*| *as normal*|, participants were presented one segment (denoted by the vertical bar) at a time by pressing the space bar or other designated key. In this case, the anaphor pronoun *he*/*she* was a target ROI. The detection of reading difficulties is manifested in longer reading times, where word recognition may be faster (i.e., less than 200 ms; Rayner et al., 2012) given there is no conflicting information but is longer for words that may contain incongruences or inconsistencies with the reader's knowledge (e.g., Heredia et al., 2016). A purpose of using a VMW paradigm was to precisely inspect the locus of the mismatch effect and determine the possibility of stereotype activation. VMW allows participants to read at their own speed and it is a more sensitive measurement of online sentence processing in comparison to the whole sentence reading methodology used in Carreiras et al. (1996).

Kennison and Trofe's (2003) paired sentences contained a gender stereotype in the first sentence and a pronoun type (*he*/*she*) in the second sentence. In the two conditions nouns were matched (e.g., *nurse ~ she*; *surgeon ~ he*) or mismatched (e.g., *nurse ~ he*; *surgeon ~ she*) in typicality of gender and anaphoric pronoun. Their results replicated the mismatch costs effects

reported by Carreiras et al., where English speakers slowed down in mismatched gender conditions in comparison to the matched conditions. However, one issue that we would like to underscore is that Kennison and Trofe's paired sentences (e.g., *The pilot announced the time and weather. She indicated that the plane would be landing a little ahead of schedule*) created a pause by the positioning of a period at the end of the first sentence. This is considered because of potential elicited processing time (e.g., wrap up effect; Raney, Campbell, & Bovee, 2014) affecting the time-course measurement of the anaphor that immediately followed making it unable to capture the processing effects at an accurate location. We avoid this potential confound by modifying their sentences into one sentence.

Kreiner et al. (2008) employed an eye-tracking methodology to examine the comprehension of gender stereotypical role nouns at sentence-level processing. Eye-movements were recorded as participants read sentences to measure how gender nouns are processed when they are presented in stereotypical (e.g., *minister*) and definitional (e.g., *king*) roles in context. The sentences in Kreiner et al.'s anaphora experiment contained an antecedent stereotypical (e.g., *minister*) or definitional (e.g., *king*) noun that was followed by a reflexive pronoun (*himself*/*herself*). Kreiner et al. looked at reading difficulties in ROIs using four main eye-movement measurements. One of the measurements was the length of time in the *first fixation* in the target NP ROI which was interpreted as reflecting an early-stage of language processing (i.e., automatic activation of the stereotyped noun). The *first-pass* time eye-movement also was interpreted as an early-stage process and was calculated as the total time that a reader's gaze fixation entered and remained in one region until exiting it. For the late-stage reading processes, one measure was the *regression-path* times, which was measured from the total time, including

times during repeated fixations, that the reader's eye would enter one word from the left to the time the eye exited the word. The *second-pass* time is the total time subtracting the times from all *first-fixations* and *first-pass* times. The *second-pass* times were included as late measures that may be times in which a reader is integrating text after the first-pass times. The sentences were divided into eight ROIs. Anaphoric sentences of the type *Yesterday, the minister left London after reminding herself/himself about the letter*) were used. ROIs were the antecedent region (article and subject noun; e.g., *the minister*), target region (reflexive pronoun; e.g., *himself*/*herself*), and spillover region (word that directly followed the pronoun; e.g., *about*).

Kreiner et al. hypothesized that the reflexive pronoun would produce reading difficulties, specifically in the target region in mismatching condition of stereotypical and definitional nouns. In cases where the stereotype of the noun mismatched its anaphor, difficulties were recorded. They also found differences between stereotypical and definitional nouns. Effects for both types of nouns were compared to inspect possible differences between nouns marked in gender (e.g., *king*) and nouns that were not marked (e.g., *doctor*) but contained a possible biased representation. In comparing definitional and stereotypical nouns, Kreiner et al. hoped to find differences that would help define whether stereotypical nouns were processed lexically or inferentially.

Their eye measurement results showed a second-pass effect in the antecedent region for the mismatch conditions presenting both stereotypical and definitional nouns. This suggested that there was no difference in processing difficulty effects for the definitional and stereotypical noun but in comparing the congruency conditions (mismatch vs. match), the recordings indicated readers consistently returned to the antecedent region (i.e., noun). Regression in eye measurements is defined as backwards movements during reading (Rayner et al., 2012). Typically, a reader's

eye- movements will return to a region that was not fully comprehended in a first-pass but may also occur for regions that become distorted to the reader upon encountering information in later text (e.g., mismatching conditions of gender agreement). Moreover, conflicting information would also prompt rereading when integration and knowledge, whether grammar or world knowledge, do not coincide (Booth & Weger, 2013; Raney & Bovee, 2016). Regressions are hypothesized to reflect processing in late stages.

Evidence for early-stage processing was found only in the antecedent region (e.g., *the minister*) and spillover region as *first-fixation* and *first-pass* measures. Given that the spillover region follows a target region, it has been considered a region where readers tend to spend additional time to process what they read in the previous region. Any longer duration fixations in this region were interpreted as reflecting the mismatch cost effects where gender agreement was violated for both noun types. These were observed in conditions where agreement in gender was difficult during stereotype processing.

In an ERP study, Osterhout et al. (1997) used reflexive pronoun and stereotypical and definitional occupational role nouns to measure differences in activation in terms of early- and late-stages of language processing. The ERP components measured were sensitive to anomalies in syntactic and semantic violations. The P600 effect is a positive voltage component that has been found to be elicited at 600 ms and is known to occur when a syntactic anomaly (e.g., *father ~ herself*) is read (e.g., Hagroot, Brown, & Groothusen, 1993; Osterhout & Holcomb, 1992). Another component that is used to detect linguistic anomalies is the N400, a negative wave that peaks at 400 ms. This ERP is thought to occur at the end of a sentence where there are semantic, or pragmatic, anomalies (Kutas & Hillyard, 1980). A pragmatic anomaly would be related to

conceptual nouns that hold stereotypical information. Osterhout et al.'s (1997) research question addressed the differences in semantic and syntactic anomalies elicited in definitional and stereotypical gender nouns. In their study, participants were presented with stereotypical violations (e.g., *bartender ~ herself*) and definitional-violations (e.g., *uncle ~ herself*) using both genders in sentences to be compared in terms of their recorded ERPs. Their experiment did not yield significant differences between stereotypical and definitional noun types in terms of an elicited mismatch cost effect, but the study indicated that socially-stereotyped gender information was sensitive to violations as recorded in the ERPs. That is, the mismatch effect reflected in the P600 was recorded for both definitional and stereotypical nouns.

Reali et al. (2015) used a German-speaking population to examine how gender is processed in a language that, like Spanish, marks gender grammatically (e.g., *Chirurg* ~ male surgeon/ *Chirurgin* ~ female surgeon) where the presence or absence of the suffix -*in* (female) cues the noun's gender morphologically in German. Reali et al. (2015) employed a priming experiment that presented descriptions and role nouns previously rated for gender typicality (male or female) to decide on either semantically related (e.g., *florist*) or unrelated (e.g., *nurse*) in order to measure if the activation of the grammatical gender was present. Reali et al. proposed the grammatical cue in a grammatical gender language (i.e., German) can be controlled for by the use of descriptors instead of nouns. In their priming experiment (Experiment 1), Reali et al. examined this by using job descriptors that referred to *bigendered* nouns (i.e., nouns that are not limited to one gender) whose suffix presence or absence specifies gender. One interest in their study was to separate grammatical gender (e.g., morphological cues) from stereotypical gender information since both had been found to be processed with interference of one another (Reali et al., 2015). This issue

had been difficult to untangle in studies that used natural gender languages such as English (e.g., Duffy & Keir, 2004; Kreiner et al., 2008), or grammatical gender languages such as German (Irmen, 2007; Reali et al., 2015) or Spanish (e.g., Carreiras et al., 1996). Overall, in experiment 1, Reali et al. found that a mismatched effect was present as revealed by faster response times for typically female descriptors (e.g., *teaches small children*) were followed by a female role noun (i.e., *lehrerin*; female teacher) vs. a male version of that noun (i.e., *lehrer*; male teacher) but not to the grammatical feminine form. However, this was not the case when typically, male descriptors (e.g., *develops software*) were presented and followed by a female noun. Both gender nouns that were related to the description in terms of occupation, showed no response time differences. This suggests that a possibility of grammatical interference may still be assumed for gender comprehension since German has grammatical gender, but masculine is the "generic" gender nonetheless (Gygax et al., 2009).

Reali et al.'s research questions addressed the theoretical issue of whether gender-stereotyped information is accessed earlier during the reading process as suggested by Kreiner et al. (2008) and Carreiras et al. (1996). That is, in using discourse presenting stereotypical nouns with mismatching anaphoric pronoun has been found to activate gender information as marked by slower reading times when participants have encountered mismatching stereotypical and grammatical information (Carreiras et al., 1996; Duffy & Keir, 2004; Kreiner et al., 2008). Reali et al. performed a second experiment using an eye-tracking paradigm to examine anaphoric resolution. Two sentences were presented as stimuli. The first sentence contained the phrase descriptor (e.g., *teaches small children*) and the second was the target sentence (e.g., *usually he/she*

has a sufficient income). As in Experiment 1, grammatical markings were omitted to control for grammatical interference of morphological markings in German role nouns.

Reali et al. divided the target sentence into five regions similar to Kreiner et al. (2008). The anaphora region included the pronoun *he*/*she* and the definite/indefinite articles (e.g., *a*, *an*, *the*) as the ROIs where the effect was expected to occur. Different eye-tracking measures were used to distinguish between early and late stages of processing gender information. The first-fixation time (early stage marker) showed longer times in the anaphor region for mismatching conditions of typically male descriptions with *she,* and shorter fixation times for matching conditions of male descriptions with *he*. These results are consistent with those from Kennison and Trofe's (2003) visual moving window, Carreiras et al.'s (1996) whole-sentence reading experiments, and the eye-tracking experiments in Kreiner et al. (2008).

Overall, results from Experiment 2 showed a mismatch effect between the subject and its anaphor in the sentences, indicating that the stereotype was activated in the early stages. Their results were interpreted under the theoretical model of world knowledge (Sanford & Garrod, 1998) where the reader comprehends through the system of probability based on previously stored information (e.g., most doctors are males).

BILINGUAL GENERAL ISSUES

Most of the studies examining gender-stereotype comprehension have looked at monolingual speakers. Likewise, the models described above were developed and modified to explain these monolingual language effects. How do bilinguals understand gender-stereotyped discourse? What are the effects of a gender-marked language (Spanish) in the comprehension of an L2 (English) not possessing such grammatical properties? In bilingualism, language dominance

is referred to as the extent to which one language is used and accessed most frequently in comparison to the other language, as well as the fluency in the L2 (Dunn & Fox Tree, 2009), is a factor that has been found to moderate relationships between comprehension and meaning activation (Heredia & Muñoz, 2015; Heredia & Cieślicka, 2016).

This study is partially motivated by our own bilingualism and bicultural experiences of gender representation (i.e., linguistically and socially). As a Spanish-English (S-E) bilingual, the knowledge of both grammatical systems in each language has facilitated our comprehension in casual conversations that switch between English and Spanish simultaneously. This code-switching is common in border city regions of South Texas due to the proximity to Mexico. One important aspect of bilingual speakers is that they are capable of adjusting their registers as they communicate with another bilingual or a monolingual, in which they mix both languages, or remain in one language. Most bilinguals in this region are exposed to both languages daily, whether by commercial signage or in conversation. The influence of Spanish and English in this region can be seen in expressions, such as *I am going to pay the light* (*electrical bill)* which are literal translations of Mexican Spanish expressions such as *Voy a pagar la luz*. This cross-linguistic interaction or *language transfer* (i.e., a linguistic phenomenon occurring when a reader applies knowledge from L1 to L2 or vice versa) is prevalent at the comprehension and production levels (see for example Kintsch & Van Dijk, 1978; Baddeley, 2000; Zwaan & Radvansky, 1998), as well as the sentential and word levels of bilingual communication. These language transfer effects have also been observed as bilinguals describe English objects, as in *I'm going to la store* (*the store ~ la tienda*). One possible explanation is that that *la* (feminine definite article) agrees in gender with the Spanish translation for *store* (*tienda*) and therefore will be processed by bilingual readers as

grammatically matched syntactic structures (Heredia & Altarriba, 2001). If bilinguals comprehend *la mesa* as having feminine gender, then *table* would be understood when processed lexically as being feminine (e.g., Heredia et al., 2016; Raney & Bovee, 2016).

Heredia et al. (2016) reported data from bilingual speakers reading mixed-language sentences of the type, *We took a walk to LA CITY before we drove back,* in which the Spanish article (*la*/*el*, *una*/*uno*) was attached to the English noun (e.g., *LA CITY*). In this case, the English noun was paired with the article that matched its Spanish translation (*LA CIUDAD*). Heredia et al. found that bilingual readers processed matching pairs faster when the English noun and the Spanish definite article matched in gender (e.g., *la mesa ~ la table*) than when there was mismatched (e.g., *el table*). Indeed, these findings support the general idea that gender agreement constraints are present for bilinguals when language switching in discourse (e.g., *la table*) is present. Given that English has conceptual gender, can influences from Spanish's grammatical gender be enough to facilitate access to gender information during comprehension?

PRESENT STUDY

The present study investigates the comprehension of gender-stereotyped information by bilingual speakers whose L1 or L2 is a gendered-marked language and operate in a bilingual community in which both languages are used frequently and code-switched during the communicative process (Heredia, Martinez, Clark, & Moreno, 2003). The aims of the current study is threefold: first, if bilinguals possess the conceptual and grammatical representation of gender-based categorization, are bilinguals more likely to exhibit a congruency mismatch cost (i.e., take longer times to comprehend) as they encounter gender-biased occupational roles that might conflict with their schematic representation that for example, *a tailor* must be a male because its

Spanish equivalent denotes a much more stereotyped gender representation in which even females are described as *un sastre* instead of the highly infrequent *una sastra*? Second, we specifically examine language dominance as a moderating factor of the mismatch cost as bilinguals encounter occupational roles conflicting with the schematic representations. Using Dunn and Fox Tree's (2009) Gradient Bilingual Dominance Scale (GBDS), bilingual speakers are classified as English-dominant, Spanish-dominant, and balanced. Thus, Spanish dominant bilinguals are expected to show significantly larger mismatching effects as compared to English dominant and balanced bilinguals. It is possible that English dominant bilinguals might be able to suppress information from their less dominant language (i.e., Spanish), thus eliciting significantly smaller mismatching effect. That is, stereotype activation and strength might very well be a function of which particular language is more readily active as the bilinguals' two memory systems interact (Heredia & Cieślicka, 2016).

A third aim is to further explore the extent to which stereotyped information is accessible during early (automatic, lexical) or late (inferential) stages of language processing. Lexical or automatic processing is hypothesized to occur early as when the first fixation lands on a stereotyped noun's anaphoric referent (e.g., *minister ~ himself*). To this end, we used eye-movements to measure processing difficulty (as revealed by reading times) as bilinguals read sentences with gender-biased occupational roles that match (*nurse ~ she*) or mismatch (*nurse ~ he*) to their mental representation.

Eye-movements have temporal value and are considered to be true on-line measures, for example, in studies that have examined implicit and covert biases. The theoretical foundation for eye-movement measurements is accounted by the eye-mind assumption, which refers to the

existing connection between eye behaviors and cognitive processes (Just & Carpenter, 1980). More specifically, ocular measurements are indirect measures of cognitive processes (i.e., controlled and automatic). The eye will remain on the word until it has been processed and comprehended by the reader before moving on to the next word. Eye-tracking techniques systematically produce a time course mapping during sentence processing, such as durations for all eye pauses on target words in an experiment (Rayner, Just, & Carpenter, 1980; Rayner et al., 2012). It is ideal to investigate anaphoric resolution in reading processes, thus prior to this method, cognitive processes in reading experiments were limited to established effects (i.e., processing difficulties) without the timeline of events (e.g., immediate activation). To elaborate briefly, a skilled reader can read at a speed of approximately 200 ms per word; this cannot be captured accurately using other measurements (e.g., moving window paradigms) to the same degree that an eye tracker can.

Conversely, experimental tasks where reaction time is measured (e.g., lexical decision tasks, VMW) instructs readers to make a judgment by choosing and pressing a button or a key. This requires the reader to process, judge, then execute a motor response before the judgment is recorded. This series of responses results in added, unknown amount of time between the text and decision, which compromises the accuracy of moment-to-moment times.

The present study follows Kreiner et al. (2008) and Reali et al.'s (2015) experimental methodology closely to address early vs. late reading processes. Experimental nouns and sentences from Kennison and Trofe (2003) were also used and further modified. Bilingual participants read sentences of the type, *The pilot* | *announced* | *that the plane was late* | *as* | *SHE/HE* | *thanked* | *passengers for flying with United*. Sentences were separated into ROIs as suggested by Reali et al.

(2015), Kreiner et al.'s (2008), and Raney et al. (2014). For example, in this sentence the antecedent is *pilot*, and the anaphor is the pronoun *she/he* so that the measurements being looked at in terms of eye-tracking fixation duration would be in the target regions where the pronoun and antecedent are. ROIs are examined through specific eye-movements that are hypothesized to detect early- and late-stage reading processes.

Stages of processing in eye-tracking measurements refer to specific points and lengths of time in which a target region and target word are read. These stages can be identified as early- or late-stages depending on the measurement that is detected and the amount of time a reader spends on any target region. Early-stages are usually reflected by the *first-fixation duration* (i.e., the first time a word is landed on and length of time spent on before moving to next region or word; Heredia & Cieslicka, 2016) and *gaze duration* (i.e., the sum of all fixations that occur before entering the next region). Early processing is considered to be an early activation when the first fixation, given a suggested amount of time, lands on a stereotyped noun's anaphoric referent (Kreiner et al., 2008). The degree to which stereotyped gender information influences context comprehension is directed toward several findings on early activation. In testing these effects, reading and eye-movements have become versatile experimental tools for interests in such thought-behavior domains.

In contrast, late stages are interpreted as the point in which the reader will be integrating previous information from representations in semantical or lexical knowledge. In research findings, late-stages are interpreted as the phase of integration or problem-solving where the reader spends the necessary time to comprehend or "correct" any mismatches within the context (Duffy & Keir, 2004; Reali et al., 2015). *Regressions* (Heredia & Cieslicka, 2016) and *right-bounded durations* reflect late-stages in processing (i.e., a sum of all the fixation durations in a single region

before the eye fixates on any region to the right; Pickering, Traxler, & Crocker, 2000). These are thought to be "late" in eye-movement measures since they are measured past the initial phase (i.e., *gaze duration*) and appear to occur after the presentation of a stimuli has been encountered.

For the early-stage processes, the measurements used in the present study included *gaze duration* and for the late-stages, *regressions*, and *right-bounded duration* times were included (Duffy & Keir, 2004; Kreiner et al., 2008; Reali et al., 2015; Heredia & Cieslicka, 2016). The percentage of skips were also used as a measurement of readers' prediction of the upcoming text and were taken as an indication of varying degrees of reading difficulties (Vasishth, Von der Malsburg, & Engelmann, 2013). That is, the higher the percentage of skips, the easier the text was processed by the reader.

METHODS

PARTICIPANTS

A total of 99 bilinguals (female *n* = 86; male n = 13) from the Texas A&M International University (TAMIU) student population participated in the study. Participants were recruited using the research participation Sona System (https://tamiu.sona-systems.com/default.aspx). Participants received class credit as partial requirement for class. Nine participants were excluded from the data due to computer errors and failing to follow directions. Participants were scheduled two at a time for a sixty-minute session. Participant's mean age was 22.6 years old (*SD* = 5.1, *range* = 15-42 years old). Participants were required to be Spanish-English or English-Spanish bilinguals. All participants had good to corrected vision. Based on their aggregated scored on Dunn and For Tree's (2009) GBDS, 28 participants were classified as "balanced" bilinguals (*M* = 55.14, *SD* = 36.48), 58 were English-dominant (*M* = 62.97, *SD* = 31.4), and 13 were Spanish dominant (*M* = 56.08, *SD* = 35.45). GBDS is a 12-item questionnaire consisting of open- and close-ended questions (e.g., *If you had to choose one language for the rest of your life, which would it be?*) that aims to determine the degree to which a bilingual is dominant in one of the two languages or if both are weighted equally (balanced bilingual). Participants' responses were scored on a range from -30 to +30 where a balanced bilingual would range between -5 and +5. English- or Spanish-weighted addresses language acquisition, everyday use of each language, attitudes toward each language, and language fluency for each. Responses were scored from this scale for the final analysis as a moderator between the target nouns and reading difficulty effects.

Table 1 summarizes the combined scores for ages that English and Spanish were acquired, years of school for English and Spanish, amounts of daily mix of languages, Spanish and English

use, and bilingual directionality (i.e., dominance). Table 2 summarizes the intercorrelations (r) between the variables that were used to measure language dominance and language proficiency. The patterns specify positive correlations between participants' daily mixing of both languages ($p < .001$) and their ability to speak and read Spanish ($p < .05$). Other positive correlations were found between proficiency variables which shows that proficiency and language usage is positively correlation in this study.

Of the simultaneous bilinguals, 26% reported learning Spanish between 0-5 years old. Of the participants that indicated English as the L1, 20% reported to having learned Spanish between 0-5 years old, 5% reported between 6-9 years old, 3% reported between 10-15 years old, and 3% reported after age 16. Of the participants that reported Spanish as their L1, 42% reported to having learned Spanish between 0-5 years old. As can be seen from Table 1, 24% of participants that indicated learning both languages at the same time (henceforth simultaneous bilinguals) reported learning English between 0-5 years old. Of these bilinguals, only 1% reported learning English between 6-9 and another 1% between 10-15 years old (see Table 3). Thirty percent of participants that indicated English as their L1 reported learning English between 0-5 and 1% learned English between 6-9 years old. Of the participants that indicated Spanish as L1, 24% reported to have learned English between 0-5 years old, 13% reported to have learned English between 6-9 years old, 4% reported 10-15 years old, and 1% after 16 years old.

An independent between-subject analysis of variance (ANOVA) was performed for the participants' age (see Table 3) and group (E-S, S-E, Simultaneous bilinguals). The analysis revealed no significant age differences between the three groups, $p > .05$. Ratings for daily

language mixing showed a significant effect, $F(2, 96) = 3.805$, $p = .026$. Multiple contrast revealed that English-Spanish (E-S) bilinguals reported less mixing than those simultaneous bilinguals.

TABLE 1. Summary of Intercorrelations for Scores of Language Variables

	1	2	3	4	5	6	7	8	9
1									
2	-.156								
3	.179	-.089							
4	-.086	.004	-.132						
5	.008	-.357	-.121	.015					
6	.273	-.418	.061	.024	.525				
7	-.324	.244	-.168	.310	-.129	-.168			
8	**-.637***	.314	-.193	.157	-.195	-.587	.451		
9	.376	.563	.233	-.065	.346	**.639***	-.427	**-.733***	

Note: 1= Age English was learned, 2 = Age Spanish was learned, 3 = Years of schooling in Spanish, 4 = Years of schooling in English, 5 = Daily mix of both languages, 6 = Daily use of Spanish, 7 = Daily use of English, 8 = English dominance, 9 = Spanish dominance; p < .05, p <01**

There was a marginal significance ($p = .07$) suggesting that Spanish-English (S-E) bilingual tended to mix (i.e., code-switch) more than E-S (see Table 4). The analysis for Spanish daily usage was reliable, $F(2, 96) = 18.19$, $p < .001$. Multiple contrasts revealed that ES bilinguals used significantly less Spanish than Bilinguals who learned both languages simultaneously.

TABLE 2. Summary of Proficiency and Language Usage

	1	2	3	4	5	6	7	8	9	10
1										
2	.053**									
3	.301*	.680**								
4	.30*	.654**	.843**							
5	.184	.541**	.708**	.719**						
6	.215*	.563**	.786**	.787**	.711**					
7	-.129	-.168	-.286*	-.254*	-.152	-306*				
8	-.107	-.159	-.173	-.186	-.090	198*	.538**			
9	-.049	-.101	-.123	-.094	-.035	.104	.435**	.800**		
10	-.049	-.016	-.104	-.111	.053	-.120	.475**	.681**	.802**	
11	-.027*	-.138	-.167	-.137	.020	-.171	.435**	.688**	.665**	.753**

Note: 1= Daily language mixing; 2= Daily Spanish usage; 3= Speaking ability in Spanish; 4 = Reading ability in Spanish; 5 = Understanding ability in Spanish; 6 = Writing ability Spanish; 7 = Daily English usage; 8 = Speaking ability in English; 9 = Reading ability in English; 10 = Understanding ability in English; 11 = Writing ability in English; *p <.05; **p < .001

S-E reported using more Spanish than E-S bilinguals. No other comparisons reach significance. The analysis for English daily usage was also reliable, $F(2, 96) = 4.99$, $p = .009$. Follow up comparisons showed that E-S rated themselves higher in using English than simultaneous bilinguals. Likewise, E-S rated themselves as using English more than S-E participants. No other effect was significant.

TABLE 3. Language Background Information for the Bilingual Sample

Variables	E-S	S-E	Simultaneous
Ages	22.9	22.6	22.3
Mean language mixing ratings	3.84	4.80	5.0
Mean English daily usage	6.687	5.95	6.0
Mean Spanish daily usage	3.22	5.33	4.69
Mean Language Proficiency Ratings			
Speaking English	6.87	6.29	6.54
Reading English	6.90	6.41	6.65
Understanding English	6.84	6.48	6.58
Writing English	6.81	6.43	6.39
Speaking Spanish	3.74	5.57	5.62
Reading Spanish	3.39	5.60	5.77
Understanding Spanish	4.52	6.24	5.58
Writing Spanish	2.74	6.54	4.73

The analysis for the years of schooling received for English were not significant (p = .57). A total of 1% of simultaneous bilingual participants reported to have no schooling in English, 4% reported receiving 1-6 years, and 21% reported 7 or more years of schooling in English. Of the participants with English and the L1, 4% reported 1-6 years of schooling in English and 27% reported having 7 or more years. Of the participants with Spanish as L1, 6% reported having 1-6 years of schooling in English and 36% reported to having 7 or more years.

The analysis for the years of schooling received for Spanish were also not significant (p = 0.88). A 7% of simultaneous bilinguals reported zero years of schooling in Spanish, 17% reported 1-6 years, and 2% reported 7 or more years. Of the English as L1 participants, 10% reported no

schooling in Spanish, 19% reported 1-6 years, and 2% reported 7 or more years. Of the participants of Spanish as L1, 10% reported no years of schooling in Spanish, 27% reported 1-6 years, and 5% reported 7 or more years.

The analysis in speaking ability in Spanish revealed a significant effect, $F(2,96) = 20.70$, $p < .001$. Multiple comparisons showed that simultaneous bilinguals rated themselves higher than E-S bilinguals. Further, S-E bilinguals rated themselves higher in speaking Spanish than E-S bilinguals. The same pattern was observed for reading Spanish, $F(2,96) = 26.03$, $p < .001$. Simultaneous bilinguals rated themselves higher than E-S bilinguals and S-E rated higher than E-S. Likewise, the analysis language understanding ability revealed a significant effect, $F(2, 96) = 16.99$, $p < .001$. Simultaneous bilingual ratings were higher than E-S, and S-E ratings were significantly higher than E-S. No other effects were significant.

The analysis by writing ability in Spanish revealed a reliable effect, $F(2, 96) = 20.25$, $p < .001$. Simultaneous bilingual ratings were higher than E-S, and S-E ratings were significantly higher than E-S. No other effects were significant. The analysis by Speaking English ability revealed a significant effect, $F(2, 96) = 5.58$, $p < .01$. Multiple comparisons showed that E-S rated themselves higher than simultaneous bilinguals. Further E-S bilinguals rated themselves higher in speaking English than S-E bilinguals.

The analysis by ability to read English was reliable, $F(2, 96) = 4.58$, $p < .001$. The only difference in reading ability was between E-S and S-E. Where E-S rated themselves higher than S-E bilinguals. The analysis by English understanding ability was not statistically reliable $F(2, 96) = 2.81$, $p = .07$, suggesting that the three groups understood English equally well. The analysis by

writing ability in English revealed a marginal effect, $F(2, 96) = 3.02$, $p = .054$. However multiple comparison revealed no differences in writing ability between the three groups.

MATERIALS

Ninety-six strongly stereotyped nouns were taken from Kennison and Trofe's (2003) normed stimuli. In Kennison and Trofe's ratings study, participants were asked to norm social role nouns on a 1-7-point Likert scale (where 1 = mostly female, and 7 = mostly male). The ranges used from Kennison and Trofe's list of rated occupational role nouns in the present study were less than 2.5 for stereotypical female occupational role nouns, and above 5.5 for stereotypical male. Once stimuli were chosen, they were organized into sets of one female (e.g., *florist*) and one male (e.g., *sheriff*) stereotyped noun to create the sentence pairs for the experiment. The sentences consisted of the two conditions, mismatching (e.g., *florist ~ he*; *sheriff ~ she*) and matching (e.g., *sheriff ~ he*; *florist ~ she*). Each condition consisted of the noun pairs that were strongly stereotyped female and strongly stereotyped male. One stereotypical gendered noun, according to ratings, was taken to form the sentence lists. In total, four sentences were constructed per list. Two of the sentences per list used a noun that was stereotypically female and the other two used a noun that was stereotypically male. Word count per sentence and the length of the word that preceded the targets were also controlled for, varying between 14-18 words per sentence. The stereotypical role noun (e.g., *florist*, *sheriff*) served as the antecedent to the anaphor (i.e., *he*/*she*). The context for the four sentences was identical except for role nouns and the pronouns, which depended on the gender stereotypicality (i.e., mostly female or mostly male) and the mismatch/match condition. Table 4 presents the four sentences in one list.

In the original sentences from Kennison and Trofe (2003), a period separated the sentence pairs, which is found to produce a wrap-up effect. The wrap-up effect has been interpreted as a reflection of increased time allowance to process the sentence (Just & Carpenter, 1980; Raney et al., 2014). To avoid a wrap-up effect, all sentences were simplified by using single sentences (e.g., *The florist stopped by the diner on the way home, and she/he ordered a catfish sandwich with fries*). Because each of the two pairs of stimuli included the exact contextual information (see Table 5), four lists were required to counterbalance the design.

TABLE 4. Sample Experimental Sentences and Conditions

Antecedent noun stereotyped to refer to mostly females	
Matching	*The florist stopped by the diner on the way home, and she ordered a catfish sandwich with fries.*
Mismatching	*The florist stopped by the diner on the way home, and he ordered a catfish sandwich with fries.*
Antecedent noun stereotyped to refer to mostly males	
Matching	*The sheriff stopped by the diner on the way home, and he ordered a catfish sandwich with fries.*
Mismatching	*The sheriff stopped by the diner on the way home, and she ordered a catfish sandwich with fries.*

A Latin Square partial counterbalancing procedure (ABCD) was utilized to assign each sentence condition to a different list, so that no experimental sentence with the same contextual information (see Table 4) was repeated within a list. Thus, stimuli assignment was between lists. A total of 48 experimental sentences were constructed. These included 24 sentences containing a female stereotypical noun and 24 containing a male stereotypical noun. Ninety-six additional

sentences were constructed to server as fillers. Filler sentences were created from Kennison and Trofe (2003) norms that were rated between 3.5 and 4.5 and were classified as neutral.

After sentences were constructed, they were separated into 7 regions of interest (ROIs). For example, *The pilot* | *announced* | *that the plane was late* | *as* | *she/he* | *thanked* | *passengers for flying with United…* Each region was identified in terms of the antecedent, *The pilot*, and anaphor *she/he*. ROIs 2, 3, and 4 were between the antecedent and anaphor and contained the verb. ROIs 6 and 7 followed the anaphor region and contained the direct object(s). The stereotypical nouns were uniformly positioned throughout the experimental sentences at ROI 1 and the anaphoric pronouns at ROI 5.

Apparatus

Sentences were arranged into the SR-Experiment Builder running on Windows OS 7 and the Eye-Link 1000 was connected to a host computer using DOS. The Experiment Builder was used to present the sentences in the order according to the counterbalanced list constructed in the pre-experiment stage. The sentences were displayed and centered in a two-line format on a 22-inch screen at 24 inches away from the keyboard below the chin rest. A white background was used to present the sentences in a black 20-inch New Times Roman font.

Eye-movements were measured using an SR-Research Eye-Link 1000 system with a sampling rate (i.e., how many times per sec the eye position is measured) of 1000Hz. Participants were monitored monocular typically on the right eye. A 9-dot formation was presented in a random order on the screen and used to calibrate fixation points to get an accurate recording of eye-movements. The calibration was validated using the same 9-dot formation. During the validation process, the Eye-Link 1000 provided the degree of visual error for each fixation point, an average

and a maximum error of all points was recorded to ensure accurate recordings during the experiment. If the average error was greater than 0.5 degrees of visual angle, the eye tracker was repositioned, and calibration process repeated. A Microsoft SideWinder Plug & Play Game Pad was used for comprehension questions responses and to move on to the next screen after they completed each trial and move on to the next sentence. Participants were asked to click on the bottom button marked 'A' for reading completion and hit the bottom triggers for true (right trigger) and false (left trigger) responses.

DESIGN

The critical eye-movements measured were *gaze duration* (sum of the duration on a target from the farthest left point of the word to the furthest right point), *regression* (sum of all fixation durations), and *right-bounded duration* (sum of all fixation durations from first fixation of a word up to the next fixation occurs on a word to the right). The design conformed to a 2 (stereotype: male vs. female) x 2 (congruency: match vs. mismatch) x 3 (language dominance: Spanish vs. English vs. balanced) mixed factorial design, with stereotype and congruency as a within-subjects factors, and language dominance as a between-subjects variable. The main manipulation was between conditions of match vs. mismatch in terms of noun-pronoun congruency. Conditions were treated as within-subjects and between-materials. The presentation of the stimuli was randomized per condition to prevent repetition and avoid suspicion of what we were measuring.

PROCEDURE

Before the experiment, all participants signed a consent form (Appendix A) upon arrival. Verbal instructions about the eye-tracker procedure and the experiment were given and

demonstrated. The participants sat at an individual cubicle where the eye-tracker and screen were positioned. The experiment ran using the Eye-Link 1000 eye-tracker and closely followed the same setup and procedures as Raney et al. (2014) and Heredia and Cieslicka (2016). During the experiment, head movements were immobilized by instructing participants to place their chin and forehead on designated areas, with elbows resting on the table, and fingers placed on the controller to be used throughout the experiment for comprehension question responses. The chinrest was 55 cm away from the screen, and the seating stool was adjusted for participant's height.

Participants were instructed to press the 'A' button on the Microsoft SideWinder to move from one sentence to the next, and to use the bottom triggers to answer 'true' (using their right index finger) or 'false' (using their left index finger) when a comprehension question (e.g., Please answer true or false: The doctor prescribed medicine) was prompted. This was done to assure that participants were paying attention to the reading material. All comprehension questions followed a filler sentence. To obtain accurate recordings, a 9-dot visual calibration was assessed before the experiment. The calibration did not always work due to the sensitivity of the camera and some participants were unable to remain still for the entire time of the experiment. All eye-movements recorded were monocular. The total calibration time took approximately 3-5 minutes per participant. Once the calibration was validated, a set of instructions that repeated the verbal instructions would appear on the screen followed by the first trial. Before each trial, a fixation point would appear on the left side of the screen to re-test accuracy to ensure the tracker would measure the sentence that followed correctly. The participant was instructed to press the 'A' button each time the fixation point was seen to continue to the next trial. Moreover, filler sentences were always followed by a true/false comprehension question. The preparation (instruction and

calibration) and experiment are estimated to take less than 60 minutes. When participants reach the end of the experiment, the last screen prompted the participant to complete a language a questionnaire on a separate cubicle. Upon completion, participants received a summary of the experiment to be read silently before exiting the lab, and they were encouraged to ask questions about the experiment and overall procedure.

RESULTS AND DISCUSSION

Participants' responses to comprehension questions were analyzed for accuracy. All participants answered the comprehension questions with an accuracy above 90%. Responses were normally distributed across the three experimental conditions. Data from four participants were excluded due to computer errors. Data from one additional participant were excluded due to an incomplete language questionnaire. Data were analyzed using linear mixed effects models (LME) using jamovi V.9.5.5, GAMLj V. 1.02 General Analyses for Linear Models module, with fixed (i.e., independent variables; pronoun congruence, stereotype, and language dominance) and random effects (i.e., items and subjects). Analyses were conducted on both early- (*gaze duration*) and late- stage (*regressions* and *right-bounded duration*) reading measures (Rayner, 1998). For all the measures, percentage of data removed, and percentage of targets skipped as a function of the experimental conditions are provided.

GAZE DURATION

A total of 10% of the data were removed because gaze durations were less than 100 ms. A Logistic Generalized Linear Model (LGLM) was used to analyze the accuracy data. The overall model accounted for 16% ($R^2 = .16$) of the variance when both the random and fixed effects were included. The overall, analysis was not statistically reliable (see Tables 5 and 6). The analysis suggests that all reading errors were normally distributed between the different experimental conditions. LME analysis on gaze duration revealed a statistical main effect of congruency where the pronoun (anaphor) matched or mismatched the gender of the occupational role, and a two-way interaction of language dominance and stereotype. No other effects reached significance (See Table 7).

TABLE 5. Log Likelihood Ratios for Gaze Duration in Anaphor Region

	X^2	df	*p*
Dominance	.777	2	.678
Stereotype	$6.5e^{-4}$	1	.980
Congruency	.429	1	.512
Dominance*Stereotype	.936	2	.626
Dominance* Congruency	.513	2	.774
Stereotype* Congruency	.005	1	.944
Dominance*Stereotype*Condition	.291	2	.865

The main effect of the stereotype match shows that participants were faster to read the pronoun (*he*/*she*) when it matched (*M* = 204.6 ms, *SE* = 6.0 ms) than when there was a mismatch (*M* = 220 ms, *SE* = 6.0 ms) between the pronoun and the noun (antecedent) describing a gendered-biased occupational role. That is, participants were faster to read the pronoun anaphor (*she*) when the stereotypical role applied to a female than to a male. These results replicated Kennison and Trofe's (2003) and Carreiras et al.'s (1996). For example, Carreiras et al.'s found longer reading times for a second sentence containing an anaphoric pronoun reference to a male stereotypical occupational role in a mismatched (e.g., *surgeon* ~ female) than in a matched condition (*surgeon* ~ male). There were also longer reading times recorded for male than female between the mismatched conditions.

The 2-way interaction of language dominance vs. stereotype (see Table 7) is described in Figure 1. As can be seen from Figure 1, participants with Spanish as the dominant language revealed longer reading times than both English dominant and balanced bilinguals for male stereotyped occupational nouns.

TABLE 6. Fixed Effects for Accuracy of Gaze Duration in Anaphor Region

				95% Confidence Interval				
Names	**Effect**	**Estimate**	**SE**	**Lower**	**Upper**	**exp(B)**	**z**	***p***
Congruency	Mismatch	.071	.11	-.141	.29	1.1	.65	.51
Stereotype	Male Typicality	.003	.11	-.211	.22	1.0	.03	.98
Dom1	E_Dom	-.093	.13	-.35	.15	.91	-.73	.47
Dom2	S_Dom	.146	.18	-.19	.53	1.16	.80	.42
Congruency* Stereotype	Mismatch* Male Typicality	.008	.11	-.21	.22	1.0	.07	.94
Congruency *Dom1	Mimatch* E_Dom	-.058	.13	-.31	.189	.943	-.46	.65
Congruency *Dom2	Mismatch* S_Dom	.130	.183	-.22	.50	1.14	.71	.48
Stereotype*Dom1	Male typicality* E_Dom	$-9.7e^{-4}$	.127	-.25	.25	.999	-.008	.99
Stereotype*Dom2	Male typicality* S_Dom	.143	.183	-.215	.512	1.154	.78	.44
Congruency * Stereotype* Dom1	Mismatch* Male typicality* E_Dom	-.032	.127	-.282	.217	.969	-.25	.80
Congruency * Stereotype* Dom2	Mismatch* Male typicality* S_Dom	.097	.183	-.263	.465	1.102	.53	.60

Follow up simple effects showed that Spanish dominant bilinguals exhibited longer reading times for male- than female occupational roles nouns), $F(1, 1258) = 4.31$, $p = .038$. These findings may provide new support for bilingual processes that occur during the interpretation of gender-biased nouns.

TABLE 7. Function of Language Dominance, Congruency, and Stereotype

	F	df 1	df 2	*P*
Dominance	1.432957	2	74.72	.2451
Stereotype	1.483037	1	1265.42	.2235
Congruency	7.740458	1	1255.37	.0055
Dominance*Stereotype	3.461568	2	1267.14	.0317
Dominance* Congruency	.569288	2	1255.86	.5661
Stereotype* Congruency	.001940	1	1255.86	.9649
Dominance*Stereotype* Congruency	1.404666	2	1257.55	.2458

One difference between the two languages spoken by bilinguals, is that Spanish marks nouns for gender (*la doctora*) and English does not (*the doctor*). As suggested, Spanish dominant bilinguals are perhaps more rigid in the classification of gender-biases in occupations perhaps because of an influence from their grammatical classification. The longer reading times may be explained by *language transfer* (i.e., in the bilingual mind, when one language's grammatical rules re-surface while processing the other; Weinreich, 1953; Odlin, 1989) phenomena that seems to produce the interference, as slower reading times, for participants that were Spanish-dominant more so than balanced bilinguals, and even more than the English-dominant bilinguals.

This is the first time that gender information processing is explored through a bilingual theoretical framework. This differs from earlier studies investigating monolinguals only and not addressing a cross-linguist or bilingual influences (Kennison & Trofe, 2003; Duffy & Keir, 2004; Kreiner et al., 2008), such as the grammatical composition and rules of other languages used. In relation to skipping the pronoun anaphor, an LGLM was performed (See Table 8). The model

produced an R^2 = .0064. The only statistically reliable effect was the 2-way interaction between match and stereotype (please see table 8).

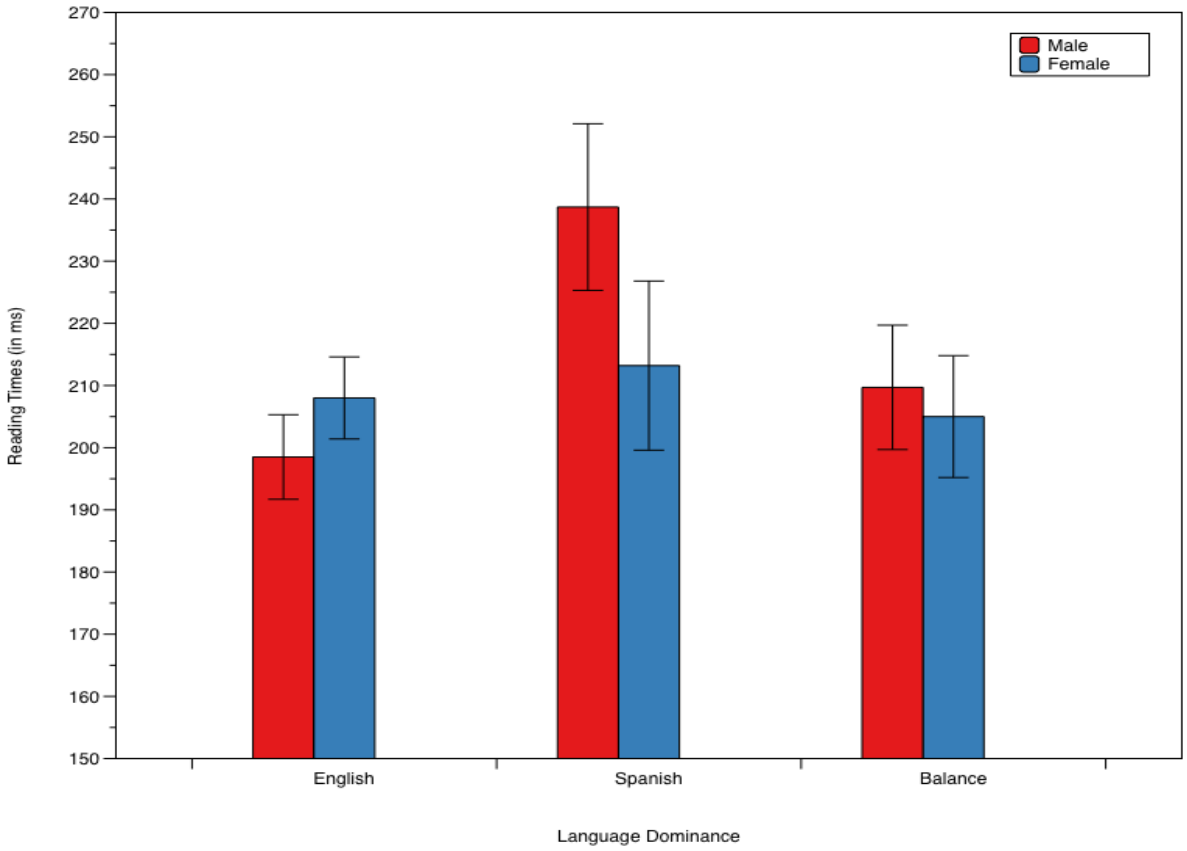

FIGURE 1. Language Dominance and Gaze Duration

Post hoc analyses using a Bonferroni adjustment show that participants were more likely to skip the pronoun anaphor when it matched the occupational role noun antecedent of males (e.g., *truck-driver ~ he*) than when it matched a stereotypically female occupation (*nurse ~ she).* Moreover, participants were more likely to skip the pronoun anaphor when it did not match a female occupation (*nurse ~ he)* than when it did (*nurse ~ she).* Skipping percentages were highest (66%) for stereotypically female nouns in a mismatched condition (e.g., *nurse ~ he*). No other effects were statistically reliable.

TABLE 8. Log Likelihood Skips for Gaze Duration in Anaphor Region

	X^2	df	*p*
Dominance	3.211	2	.201
Stereotype	.589	1	.443
Congruency	1.856	1	.173
Dominance*Stereotype	1.22	2	.543
Dominance* Congruency	.234	2	.890
Stereotype* Congruency	14.055	1	< .001
Dominance*Stereotype*Condition	.281	2	.869

Table 9 depicts the skipping amounts percentage per condition and gender stereotypicality. Skipping refers to an area that has a reader jumps when no time is needed to process any items in the specific region. The skips were measured for the anaphor region (*he*/*she*) and are interpreted as indicating early-stage processing for the stereotypical nouns matched in anaphor (vs. mismatched and vs. female mismatch).

Interestingly, overall skipping percentages for the male stereotyped noun in a matched condition (*he*) is comprehended with little to no effort, which was expected when looking at the comparison to mismatched conditions, where less skips would have been predicted. However, when comparing the differences between conditions for female and male stereotypical nouns, the skipping percentages were the opposite. That is, while stereotypical male nouns in matched conditions were skipped more than mismatched conditions, the stereotypical female nouns in matched condition were skipped less than the mismatched condition. There are two possible reasons to account for the differences between the gender bias effects. One is that the influence from concepts are represented generically as masculine in Spanish (Gygax et al., 2009). That although the stereotype is stronger toward a feminine role in English, the Spanish language rule

expects a more explicit specification on the gender, such as a definite article (e.g., *el*/*la*) or a morphological cue (e.g., *-a*, *-o*). This possibility would support the lexical representation of gender, at least for bilinguals that are Spanish dominant. The other possibility is that while American culture is progressive, and women have been in the workforce for over half a century (Clancy, 2014), the stereotype of a female profession is implicitly present almost as an oxymoron. If this is the case, it will support an inferential model more so than the lexical view. However, more in-depth analyses would be necessary to distinguish between whether the skips are occurring due to influences of language dominance or influences from implicit world knowledge.

TABLE 9. Gaze Duration Skips in the Anaphor Region

				95% Confidence Interval	
Congruency	Stereotype	Mean	SE	Lower	Upper
Match	F	.56	.019	.527	.601
Mismatch	F	.66	.018	.623	.694
Match	M	.65	.018	.613	.684
Mismatch	M	.61	.019	.568	.641

FIRST-PASS REGRESSION

An LGLM analysis was performed on the percentage of first-pass regressions as a function of language dominance, stereotype, and congruency. The overall model accounted for about 2.5% of the variance ($R^2 = .025$). See Table 10 and 11 for a summary of the results. Language dominance was the only effect that reached significance. Follow up multiple comparisons revealed that Spanish dominant (*M* = .55, *SE* = .023) participants produce more statistically significant regressive fixations to the antecedent noun than English dominant (*M* = .31, *SE* = .010) and balanced bilinguals (*M* = .28, *SE* = .015). English dominant bilinguals produced more regressions

than balanced bilinguals. This is indicative of dominant bilinguals "verifying" that the antecedent matched the pronoun anaphor. This effect might be due to how gender is represented in Spanish.

A Spanish dominant bilingual will be more likely to predict the upcoming pronoun by referring to their mental lexicon in Spanish. For example, reading the antecedent *surgeon* would produce the expectation of *he* because in Spanish surgeon is represented as masculine (*el cirujano*) and therefore, only increase the expectation beyond a stereotype. Although congruency did not interact with dominance, we can speculate that Spanish dominant bilinguals were more likely to regress in order to "recheck" the occupational role noun (antecedent region) which was interpreted as indicating reading difficulties at encountering the mismatching pronoun. English dominant bilinguals, on the other hand, were more likely to skip the antecedent.

Another interesting finding was that this effect happened for the male typical and less for the female typical nouns. This evidence supports the results from gaze duration in which Spanish dominant bilinguals took longer to read the anaphor in sentences that contained a mismatched pronoun and a male stereotypical noun but took longer to read the matched pronoun in sentences that contained a female stereotypical noun.

TABLE 10. Log Likelihood for First-Pass Regressions in Antecedent Region

	X^2	df	*P*
Dominance	106.21	2	< .001
Stereotype	.075	1	.784
Congruency	.648	1	.421
Dominance*Stereotype	.713	2	.700
Dominance* Congruency	.271	2	.873
Stereotype* Congruency	2.641	1	.104
Dominance*Stereotype*Condition	1.129	2	.569

TABLE 11. Log Likelihood for Skips in the Antecedent Region

	X^2	df	*p*
Dominance	12.35	2	.002
Stereotype	.54	1	.463
Congruency	.34	1	.559
Dominance*Stereotype	.212	2	.900
Dominance* Congruency	.604	2	.740
Stereotype* Congruency	1.64	1	.201
Dominance*Stereotype* Congruency	2.42	2	.299

SKIP PERCENTAGES

Data was also analyzed in relation to the proportion of skipping the antecedent. The results are summarized in Table 12. The overall model produced an R^2 = .0057. Like the first-pass regression analysis, the only reliable effect was language dominance. Follow up post hoc analysis revealed that balanced English dominant bilinguals (*M* = .134, *SE* = .007) were more likely to skip the antecedent than balanced (*M* = .099, *SE* = .009) and Spanish dominant bilinguals (*M* = .089, *SE* = .013). The decreased skip percentage in comparison to the English dominant and balanced bilinguals in this region indicates that the processing speed decreased for the Spanish dominant bilinguals.

ERROR ANALYSIS

Thirty-four percent of the data were excluded due to errors. An LGLM was performed on error rate. The model produced an R^2 = .0056. The results are summarized in Table 13. There was a significant interaction between congruency and stereotype. Please see Figure 2. Post hoc multiple comparisons using a Bonferroni adjustment reveal that participants made less errors when the

anaphor matched the male occupational (*M* = .32, *SE* = .018), than when it matched the female occupational (*M* = .39, *SE* = .019). Moreover, participants made less errors when there was a mismatch between stereotype and female (*M* = .31, *SE* = .018) than when female matched the stereotype which further supports the predicted mismatch effect. Figure 2 depicts the differences in error totals made between the conditions of congruency and of gender stereotype. This is interpreted similar to the skip percentage from the gaze duration analysis. The errors made are inferred as processing difficulties as readers encountered the conditions that were marked in a higher amount of error.

READING DURATION

LME analysis on the *right-bounded duration* produced an R^2 = .173 with the combined fixed and random factors. The reading time means are summarized in Table 14 and Figure 3. There was a main effect of correspondence showing a mismatch effect. That is, matching (*M* = 218 ms, *SE* = 7.25 ms) anaphors to the antecedent were read faster than mismatching anaphors (*M* = 234 ms, *SE* = 7.1 ms). This mismatch effect found for the *right-bounded duration* measure is compatible and supports the same effect found in the *gaze duration*. This shows the same mismatch effect of processing difficulties in the anaphor region.

The interaction between language dominance and stereotype shows a pattern in which Spanish dominant bilinguals took longer to read male stereotypes than female. See Table 14 and Figure 2. However, the differences did not reach significance, $F(1,1154.5) = 3.33$, $p = .07$. The right-bounded duration was assessed in order to identify if any late-stage processing were occurring for the anaphor regions. This effect typically happens when the reader fixates on a region and then moves on to a next region to the right. Unless that next region is fixated on, it is assumed

that the reader is still processing the last region where fixation last happened. The time the right-

TABLE 12. Fixed Effects for First-Pass Regression in Antecedent Region

				95% Confidence Interval				
Names	**Effect**	**Estimate**	**SE**	**Lower**	**Upper**	**exp(B)**	**Z**	***p***
Congruency	Mismatch	-.034	.042	-.117	.049	.967	-.805	.421
Stereotype	Male Typicality	-.012	.042	-.094	.071	.988	-.27	.784
Dom1	E_Dom	-.2622	.051	-.362	-.163	.769	-5.18	<.001
Dom2	S_Dom	.6989	.068	.566	.833	2.01	10.28	<.001
Congruency* Stereotype	Mismatch* Male Typicality	-.0686	.042	-152	.014	.934	-1.62	.104
Congruency *Dom1	Mimatch* E_Dom	-.0208	.051	-.120	.079	.979	-.41	.682
Congruency *Dom2	Mismatch* S_Dom	-.0037	.068	-.137	.130	.996	-.06	.956
Stereotype*Dom1	Male typicality* E_Dom	-.0236	.051	.123	.076	.977	-.47	.641
Stereotype*Dom2	Male typicality* S_Dom	.0573	.068	-.076	.191	1.06	.84	.399
Congruency * Stereotype* Dom1	Mismatch* Male typicality* E_Dom	.0319	.051	-.067	.131	1.03	.63	.529
Congruency * Stereotype* Dom2	Mismatch* Male typicality* S_Dom	-.0719	.068	-.205	.061	.931	-1.06	.290

bounded duration is summed up is the time from the first time a region is fixated until the time it takes for a next fixation to happen, granted the next fixation is to the right of that region.

TABLE 13. Log Likelihood for Right-Bounded Duration in Anaphor Region

	X^2	df	*p*
Dominance	3.22	2	.200
Stereotype	.52	1	.469
Congruency	1.19	1	.276
Dominance*Stereotype	1.39	2	.498
Dominance* Congruency	.134	2	.935
Stereotype* Congruency	11.75	1	<.001
Dominance*Stereotype* Congruency	.098	2	.952

TABLE 14. Means for Right-Bounded Duration

				95% Confidence Interval	
Congruency	Stereotype	Means	SE	Lower	Upper
Match	F	.392	.019	.357	.430
Mismatch	F	.310	.018	.276	.346
Match	M	.316	.018	.283	.352
Mismatch	M	.358	.018	.323	.395

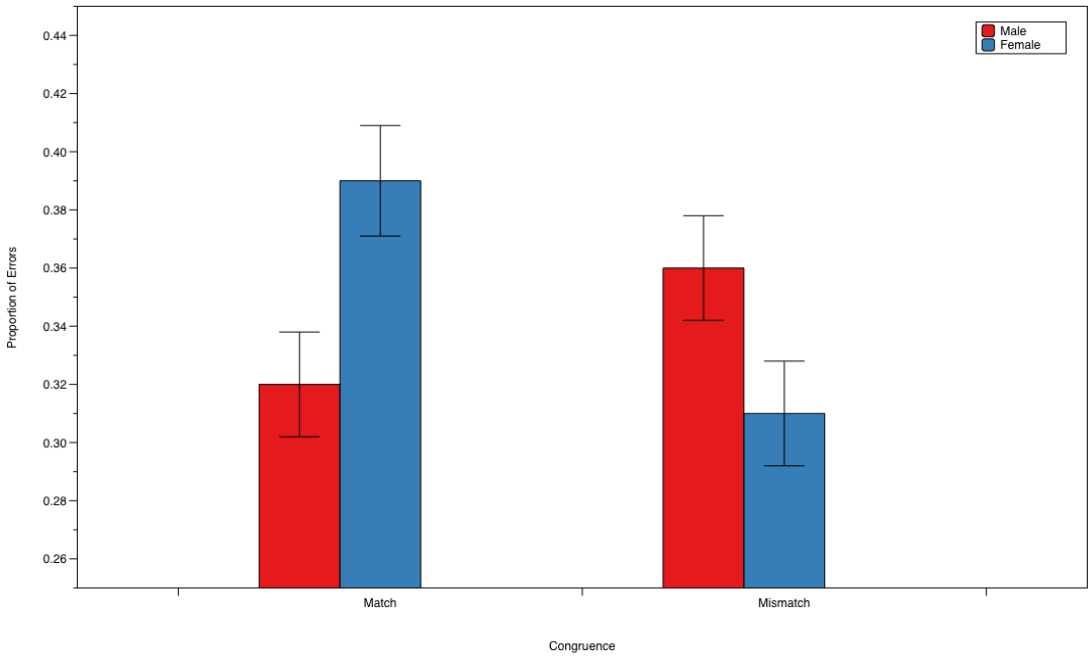

FIGURE 2. Right-Bounded Duration Errors between Congruency and Stereotype.

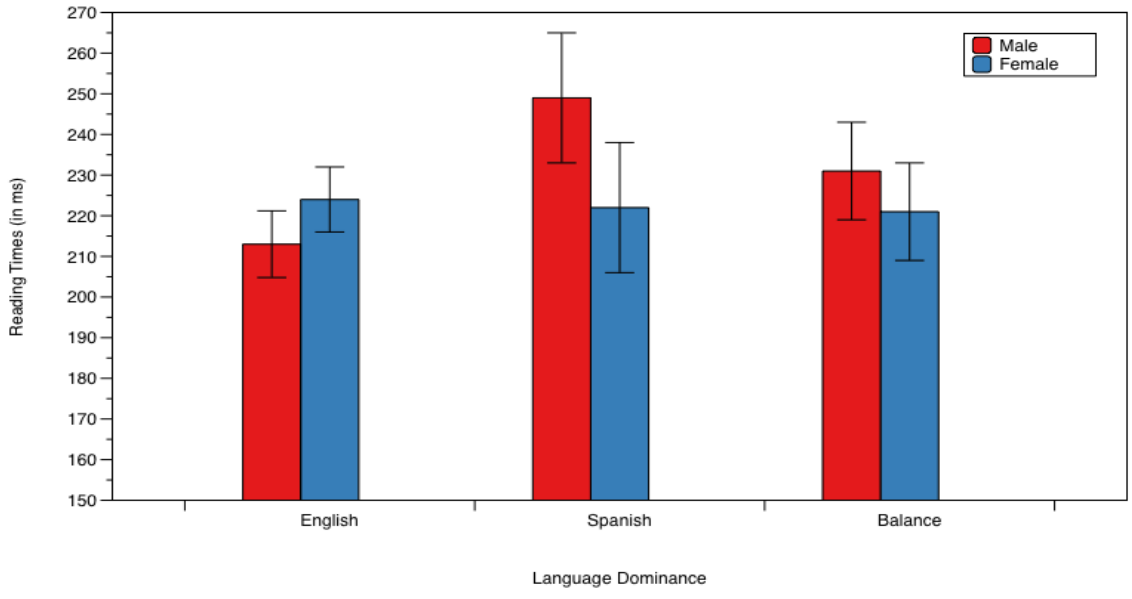

FIGURE 3. Right-Bounded Duration Reading Times.

TABLE 15. Fixed Effects for Right-Bounded Duration in Anaphor Region

				95% Confidence Interval				
Names	**Effect**	**Estimate**	**SE**	**Lower**	**Upper**	**df**	**t**	***P***
Congruency	Mismatch	7.15	3.50	.282	14.01	265.8	2.04	.042
Stereotype	Male Typicality	4.31	3.52	-2.60	11.22	259.2	1.223	.223
Dom1	E_Dom	-8.04	7.45	-22.63	6.56	75.2	-1.08	.284
Dom2	S_Dom	8.79	10.26	-11.31	28.89	72.8	.857	.394
Congruency* Stereotype	Mismatch* Male Typicality	3.36	3.50	-3.50	10.23	266.9	.960	.338
Congruency *Dom1	Mimatch* E_Dom	-2.91	3.89	-10.53	4.71	1257.7	-.748	.455
Congruency *Dom2	Mismatch* S_Dom	2.03	5.30	-8.36	12.43	1209.2	.383	.702
Stereotype*Dom1	Male typicality* E_Dom	-9.66	3.91	-17.32	-2.00	1267.4	-2.47	.014
Stereotype*Dom2	Male typicality* S_Dom	9.17	5.32	-1.26	19.61	1212.2	1.72	.085
Congruency * Stereotype* Dom1	Mismatch* Male typicality* E_Dom	.552	3.89	-7.07	8.18	1260.5	.142	.887
Congruency * Stereotype* Dom2	Mismatch* Male typicality* S_Dom	8.14	5.30	-2.25	18.53	1208.6	1.54	.125

GENERAL DISCUSSION

The purpose of the present this study was to further extend research that has looked at the comprehension of gender stereotypes by specifically looking at language dominance and bilinguals. The investigations in this study were threefold. First, participants were divided into three groups, (i.e., Spanish dominant, English dominant, and balanced bilingual) given the extent of having knowledge and using another language (i.e., Spanish) that typically marks concepts with grammatical gender were assessed for reading times and language dominance. The scores for language dominance were included to compare the reading behaviors (i.e., processing) in the different groups. Second, language dominance scores were also measured as a factor that influences the schematic representation of an occupational role noun regarding gender. The sentences were divided into ROIs to specifically look at the anaphor and the antecedent regions and were measured through specific eye measurements (*gaze duration*, *first-pass regression, right-bounded duration*) that fulfilled the third aim of this study, which was to answer the question of whether bilinguals processed gender information at the early- or late-stages of processing. It was predicted that the reading difficulties (i.e., slower reading times) would occur more in the mismatched congruency conditions (i.e., larger amounts of regressions on the antecedent, gaze durations on anaphor region and less skipping percentages).

Four main models were used as theoretical frameworks for this study. The first one was a *mental model*, which explains that a reader used previously learned information to inference gender information. The second model was the *schema*, which is similar in that it is a representation of an overall view of a concept. For example, if a person encounters mostly male doctors, it is likely they would draw from this experience and assume that a doctor is male and

would less likely be female. A third model was the *inferential view*, where this schematic representation is assumed to create probabilistic algorithms that are based on world knowledge. Lastly, the *lexical vie*w is also looked at as a possible way that bilinguals access gender information in language. In this view, readers are theorized to access gender through a mental lexicon where they make sense of gender information by a rigid language system that is made up of syntactic rules.

Gaze duration was analyzed for the anaphor region. Gaze duration is interpreted as early-stage processing because it measures how long a reader needs to remain on a critical region before moving on. Moving on indicates the region has been processed and understood, therefore this measure is interpreted as a time taking before comprehending a region. Spanish dominant bilinguals showed an increase of gaze duration in the anaphor regions in comparison to the English dominant and balanced bilinguals. In this study, *gaze duration* showed an effect for Spanish dominant bilinguals. This shows that for these bilinguals, the Spanish language does influence the processing speed and overall comprehension upon encountering a stereotypical noun and a mismatching anaphor. Since Spanish is marked in grammatical gender, bilinguals with dominance in it will expect an upcoming gender confirmation. In the mismatching congruency conditions, the stereotypical noun *surgeon* would be read and expected to be *he* according to their concept of the Spanish translation of surgeon (i.e., *el cirujano*). This expectation was measured through the gaze duration of the anaphor, where Spanish dominant bilinguals took a significantly longer time to read a mismatching female pronoun (*she*) for sentences containing a stereotypically male noun (e.g., *surgeon*).

Skipping percentages in the anaphor region also differed in processing speeds, however these were only found for the gender of the stereotyped noun. In sentences that contained a stereotypically female noun, all three groups of readers skipped less in comparison to when they read a stereotypically male noun. This suggested that Spanish dominant readers slowed down at the gender pronounce (*he* or *she*) pronoun in the mismatched conditions and could be explained once again as an influence of their dominant language's grammatical gender system. This was an interesting finding because the differences were found between female and male noun stereotype and were opposite in terms of congruency conditions. The skipping percentages were inferred as early-stages and were found in the overall analysis across the three bilingual groups. The result of skips in the anaphor region are explained as a cultural effect where a female in the workforce is not processed as quick as a male in the workforce would be.

Skipping percentages were also analyzed in the antecedent region as well. Language dominance was the only variable that showed an effect. As in the gaze duration, Spanish dominant bilinguals showed a lower percentage of skips in comparison to the English and balanced bilinguals. This means that Spanish dominant bilinguals spent more time in the antecedent region in comparison to the other two bilingual groups. One possible explanation is that Spanish influences the processing speed, given that Spanish dominant bilinguals are less likely to skip the antecedent region. This result also showed a difference in the gender of the stereotypical noun, similar to the skip percentages in the anaphor region. Stereotypically female nouns were skipped less in the matched than in the mismatched conditions and the male nouns were skipped more in the matched than in the mismatched conditions. Interestingly, it was predicted in the study that the

matched conditions would show a higher overall skipping percentage due to the processing difficulties when encountering a mismatch between noun and anaphor.

Spanish dominant bilingual readers also showed an effect for first-pass regression in the antecedent regions. This regression measure is a late-measure that is interpreted as processing and solving a reading difficulty by "rechecking" the previously read noun in the antecedent region. This would be explained by an inferential view of comprehension since Spanish is marked in grammatical gender, thus, when a bilingual reader is dominant in Spanish the applied strategies when a conceptual noun such as *surgeon* is read, they will automatically refer to their dominant language where *surgeon* is claimed as masculine (*el cirujano*).

The *right-bounded duration* in the anaphor region showed that participants would process the male and female stereotype mismatch condition for a longer time, supporting the mismatch effect predicted. In comparing the female and male stereotypical nouns in each congruency condition, the right-bounded duration was longer than when the noun was stereotypical male than in the female in the mismatch condition. This means that when the noun was stereotypically male was presented prior to a *she*, the pronoun would be processed for a longer time than if the pronoun presented was *he* and more so than either of congruency condition of the female stereotype noun sentences. This fixation is considered a late-stage measure since it is processed past the critical target region, in this case, the antecedent. The *right-bounded duration* in the anaphor region is different from the *regression* measure in that it measures the time it takes to get another fixation to the right and the *regression* is measured as the eye moves leftward. However, in this study both measurements reflect processing difficulties for the anaphor pronoun when it does not match the antecedent. In addition, the accuracy analysis for the *right-bounded duration* produced results for

the errors that readers made during the mismatched conditions. These results were compatible with the results from the gaze duration analysis for skip percentages that showed differences in the gender of the stereotype gender of nouns. Taken together, these results in the *right-bounded duration* support the mismatch effect by inferring that reading difficulties increased in sentences containing a male noun more than in those containing a female noun.

Results from the *gaze durations* in the anaphor region and skip percentages in the anaphor and antecedent regions were taken as measuring early-stages of reading comprehension hypothesized to be automatic and lexically driven. Eye movements known to measure late reading processes were also used. *First-pass regression* eye movements were employed to measure late integrative reading processes. This eye movement was significant for Spanish dominant bilinguals that exhibited significantly more regressions to the antecedent (the occupational noun) than English dominant and balanced bilingual. This suggested that Spanish dominant bilinguals might be accessing gender-stereotype information using their knowledge from Spanish to infer the gender of an antecedent prior to encountering a pronoun anaphor that would otherwise resolve the ambiguity. Another late measure used was the *right-bounded duration* for the anaphor region, which indicated a time difference between congruency conditions and the gender of the stereotype. The early and late events from these results support an inferential view of gender comprehension. The mismatch cost effects were shown in *gaze duration* times for the anaphor region. Overall, the results showed processing speed differences between the bilingual groups and for both congruency conditions. The eye-movement measures indicated that readers would need to slow down (e.g., in *gaze duration*, *regressions*, *right-bounded durations*) or would be able to speed up (e.g., skipping). With the varying differences per congruent condition and ROI (anaphor and antecedent), it was

clear that the mismatch effect was present and stereotypical gender does pose difficulties in comprehension when the stereotype is not matched between noun and pronoun. This was seen more in the Spanish dominant bilinguals, which confirms the hypothesis that knowing a language that has grammatical gender will most likely influence a readers' expectation (e.g., Heredia & Cieslicka, 2016), even when the noun is conceptually in English or the conversation is in English.

Overall, the results found support for previous literature (Carreiras et al., 1996; Osterhout et al., 1997; Kennison & Trofe, 2003; Duffy & Keir, 2004; Reali et al., 2015). The findings extend on the mismatch costs found the previous studies that measured stereotypical gender processes by introducing a bilingual dominance measure. To the best of our knowledge, this is the first time that language dominance is assessed and included in the analysis of eye-movement behaviors during gender-stereotype language processing.

Limitations

Limitations encountered were the number of participants classified as Spanish dominant that were disproportionate to the English dominant and balanced bilinguals. By increasing the number of participants in future studies, the proportion of dominance might be better distributed in order to verify that the effects reported in the present study were indeed due to language dominance with more confidence. Another limitation was that the nouns used for this study were taken from Kennison and Trofe (2003) where participants from Oklahoma State University (OSU) rated the nouns as mostly male to mostly female. The populations of the university from Oklahoma and from South Texas are not culturally alike. In fact, in 2003, OSU the university had only a 1.9% Hispanic enrollment, as compared to a 25.3% reported in 2016 (Institutional Research and Information Management, 2005; 2016). Aside from cultural differences, the time between when

their nouns were published and when they were used in this study was approximately 14 years. Future work in this area would benefit from norming Kennison and Trofe's (2003) with a South Texas population of Hispanic-Americans. Another issue with the present study was that a lexical view was not fully discounted as having no support because the nouns were conceptual and not definitional. If both noun types (i.e., stereotypical and definitional nouns) were used in this study, the comparison would determine whether readers processed the nouns lexically or inferentially. Adding a definitional noun to an additional list of sentences, would expand the present study and possibly produce a clearer indication of how gender is comprehended by bilinguals.

Language is embedded in thought which transfers to everyday interactions inevitably. One possible and practical implication is not an immediate one, but rather one that can be an addition to the movement of un-gendering language for the purpose of balancing gender representation. If it can be shown that language is influential, as recorded in microscopic mental events, a revision to how gender is used in language might be important for advocates of progressivism that seek an increase in overall balance.

REFERENCES

Allport, G.W. (1954). *The Nature of Prejudice*. Cambridge, Mass: Addison-Wesley Pub. Co.

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Science*, *4*, 417-423.

Banaji, M. R., & Hardin, C. D. (1996). Automatic stereotyping. *Psychological Science*, *7*, 136–141.

Bargh, J. (1994). The four horsemen of automaticity: Awareness, efficiency, intentions and control. In R. Wyer & T. Srull (Eds.), *Handbook of Social Cognition* (pp. 1-40). Mahwah, NJ: Lawrence Erlbaum Associates Inc.

Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology*. Cambridge, UK: Cambridge University Press.

Bobrow, R. J. & Brown, J. S. (1975). Systematic understanding – synthesis, analysis and contingent knowledge in specialized understanding systems. In Bobrow, D. G. & Colins A. (Eds.), *Representation and Understanding: Studies in Cognitive Science* (pp. 103 - 129). New York: Academic Press.

Booth R. W. & Weger U. W. (2013). The function of regressions in reading: Backward eye-movements allow rereading. *Memory & Cognition*, *41*, 82-97.

Cain K. & Oakhill, J.V. (1999). Inference making ability and its relation to comprehension failure in young children. *Reading and Writing*, *11*, 489-503.

Carreiras, M., Garnham, A., Oakhill, J., & Cain, K. (1996). The use of stereotypical gender information in constructing a mental model: Evidence from English and Spanish. *Quarterly Journal of Experimental Psychology*, *49*, 639 - 663.

Cieslicka, A. B. & Heredia, R. R. (2016). Priming and online multiple language activation. In: Heredia R. R. & Cieslicka, A. B. (Eds.), *The Bilingual Mind and Brain Series, 1, Methods in Bilingual Reading Comprehension Research* (pp. 123–156). New York: Springer.

Clancy, S. (2014). The hidden reason women aren't making it to the top. Retrieved from https://www.forbes.com/sites/85broads/2014/03/31/the-hidden-reason-women-arent-making-it-to-the-top/

Deaux, K. (1995). How basic can you be? The evolution of research on gender stereotypes. *Journal of Social Issues, 51*, 11-20.

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology, 56*, 5-18.

Duffy, S. A., & Keir, J. A. (2004). Violating stereotypes: Eye-movements and comprehension processes when text conflicts with world knowledge. *Memory & Cognition, 32*, 551–559.

Dunn, A. L., & Fox Tree, J. E. (2009). A quick, gradient Bilingual Dominance Scale. *Bilingualism: Language and Cognition, 12*, 273–289.

Fiske, S. T., Cuddy, A., Glick, P. (2007). Universal dimensions of social cognition: warmth and competence. *Trends in Cognitive Sciences, 11*, 77-83.

Forbes, J. N., Poulin-Dubois, D., Rivero, M. R., Sera, M. D. (2008). Grammatical gender affects bilinguals' conceptual gender: Implications for linguistic relativity and decision making. *The Open Applied Linguistics Journal, 1*, 68-76.

Garnham, A. (1981). Anaphoric reference to instances, instantiated and non-instantiated categories: A reading time study. *British Journal of Psychology, 72*, 377-384.

Garnham, A. (2001). *Mental models and the interpretation of anaphora*. Hove: Psychology Press.

Garrod S. C. & Sanford A. J. (1982). The mental representation of discourse in a focused memory system: Implications for the interpretation of anaphoric noun phrases. *Journal of Semantics 1*, 21-41.

Garrod, S. C. & Sanford A. J. (1994). Resolving sentences in a discourse context: How discourse representation affects language understanding. In M. Gernsbacher (Ed.) *Handbook of Psycholinguistics* (pp. 699-717). New York: Academic Press.

Greene, S. B., McKoon, G., & Ratcliff, R. (1992). Pronoun resolution and discourse models. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *18*, 266-283.

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: attitudes, self-esteem, and stereotypes. *Psychological Review, 102,* 4–27.

Gygax, P., Gabriel, U., Sarrasin, O., Oakhill, J., & Garnham, A. (2009). Some grammatical rules are more difficult than others: The case of the generic interpretation of the masculine. *European Journal of Psychology of Education*, *24*, 235.

Hagroot, P., Brown, C, & Groothusen, J. (1993). The syntactic positive shift as an ERP measure of syntactic processing. *Language & Cognitive Processes, 8*, 337-364.

Heilman, M., & Eagly, A. H. (2008). Gender stereotypes are alive, well, and busy producing workplace discrimination. *Industrial and Organizational Psychology*, *1*, 393-398.

Heredia, R. R. & Altarriba, J. (2001). Bilingual language mixing: Why do bilinguals code-switch? *Current Directions in Psychological Science*, *10*, 164-168.

Heredia, R. R. & Cieslicka, A. B. (2016). Metaphoric reference: An eye-movement analysis of Spanish-English and English-Spanish bilingual readers. *Frontiers in Psychology*, *7*, 439.

Heredia, R. R., Cieślicka, A. B., & Altarriba, J. (2016). Introduction to bilingual research. In: Heredia R., Altarriba J., & Cieślicka A. (Eds), *Methods in Bilingual Reading Comprehension Research. The Bilingual Mind and Brain Book Series* (Vol. 1, pp. 1-9). New York: Springer.

Heredia, R. R., Lopez, B., Garcia, O. Altamira, W. & Gonzalez, P. (2016). Bilingual reading: The visual moving window. In R. R. Heredia, J. Altarriba, & A. B. Cieślicka (Eds.). *Methods in bilingual reading comprehension research,* (pp. 99-121). New York: Springer.

Heredia, R. R., & Muñoz, M. E. (2015). Metaphoric reference: A real-time analysis. In R. R. Heredia & A. B. Cieślicka (Eds.), *Bilingual figurative language processing* (pp. 89–116). New York: Cambridge Press.

Heredia, R. R., Martínez, K. M., Clark, D., & Moreno, M. (2003). Processing code-switched sentences in Spanish-English bilinguals: Context and word frequency effects. Paper presented at the XIII Congress of the European Society for Cognitive Psychology, Granada, Spain.

Hess, D. J., Foss, D. J., & Carroll, P. (1995). Effects of global and local context on lexical processing during language comprehension. *Journal of Experimental Psychology: General*, *124*, 62-82.

Hillert D., & Nakano, Y. (2016). Second language sentence processing: Psycholinguistic and neurobiological research paradigms. In R. R. Heredia, J. Altarriba, & A. B. Cieślicka (Eds.), *Bilingual Reading Comprehension Research*, (pp. 231-263). New York: Springer.

Institutional Research and Information Management. (n.d.). Retrieved from https://irim.okstate.edu/

Irmen, L. (2007). What's in a (role) name? Formal and conceptual aspects of comprehending personal nouns. *Journal of Psycholinguistic Research, 36,* 431-456.

Johnson-Laird, P. N. (1983). *Mental Models: Towards a Cognitive Science of Language, Inference, and Consciousness*. Harvard University Press, Cambridge, MA.

Johnson-Laird, P. N. (2010). Mental models and human reasoning. *Proceedings of the National Academy of Sciences, 107,* 18243-18250.

Jurafsky, D. (1996). A probabilistic model of lexical and syntactic access and disambiguation. *Cognitive Science, 20,* 137-194.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review, 87,* 329–354.

Katz, D. & Braly K. W. (1935). Racial prejudice and racial stereotypes. *The Journal of Abnormal and Social Psychology, 30,* 175-193.

Kazanina, N., Lau, E. F., Lieberman, M., Yoshida, M., & Phillips, C. (2007). The effect of syntactic constraints on the processing of backwards anaphora. *Journal of Memory and Language, 56,* 384-409.

Kennison, S. M., & Trofe, J. L. (2003). Comprehending pronouns: A role for word-specific gender stereotype information. *Journal of Psycholinguistic Research, 32*, 355-378.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review, 95*, 163-182.

Kintsch, W. & Van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, *85*, 363-394.

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.

Kurby, C. A., Britt, M. A., & Magliano, J. P. (2005). The role of top-down and bottom-up processing between-text integration. *Reading Psychology*, *26*, 335-362

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

Kutas, M. & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, *207*, 203–208.

Lauro, J. & Schwartz, A. I. (2018). Cognate Effects on Anaphor Processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. Advance online publication. http://dx.doi.org/10.1037/xlm0000601.

Marslen-Wilson, W. D. & Welsh. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology*, *10*, 29-63.

Mason R. A. & Just, M. A. (2007). Lexical ambiguity in sentence comprehension. *Brain Research*, *1146*, 115-127.

McCormick, W. M. (1988). *Theories of reading in dialogue: An interdisciplinary study*. New York: University Press of America.

McKoon, G. & Ratcliff, R. (1992). Inference during reading. *Psychological Review*, *99*, 440-466.

Oakhill, J., Garnham, A., Reynolds, D. (2005). Immediate activation of stereotypical gender information. *Memory and Cognition*, *33*, 972–83.

Odlin, T. (1989). *Language transfer: Cross-linguistic in language learning*. Cambridge University Press.

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

Osterhout, L., Bersick, M., McLaughlin, J. (1997). Brain potentials reflect violations of gender stereotypes. *Memory & Cognition*, *25*, 273–85.

Pickering, M., Traxler, M. J. & Crocker, M. W. (2000). Ambiguity resolution in sentence processing: Evidence against frequency-based accounts. *Journal of Memory and Language*, *43*, 477-475.

Raney, G. E., & Bovee, J. C. (2016). Reading integration in bilingual speakers. In: Heredia R., Altarriba J., Cieślicka A. (Eds.) *Methods in Bilingual Reading Comprehension Research. The Bilingual Mind and Brain Book Series*, (Vol. 1, pp. 157-182). New York: Springer.

Raney, G. E., Campbell, S. J., & Bovee, J. C. (2014). Using eye-movements to evaluate the cognitive processes involved in text comprehension. *Journal of Visualized Experiments, 83*, 50780.

Rappaport M., Levin B., Laughren M. (1993). Levels of Lexical Representation. In: Pustejovsky J. (Eds.), *Semantics and the Lexicon. Studies in Linguistics and Philosophy* (Vol. 49, pp. 37-54). Dordrecht: Springer.

Rayner, K. (1998). Eye-movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*, 372–422.

Rayner, K., Pollatsek, A., Ashby, J., & Clifton, C., Jr. (2012). *Psychology of reading*. New York: Psychology Press.

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.

Reynolds, D., Garnham, A., & Oakhill, J. V. (in press). Evidence of immediate activation of gender information from a social role name. *Quarterly Journal of Experimental Psychology: A.*

Sanford, A. J. & Garrod, S. C. (1989). What, when, and how?: Questions of immediacy in anaphoric reference resolution. *Language and Cognitive Processes*, *4*, 235-263.

Sanford, A. J. & Garrod, S. C. (1998). The role of scenario mapping in text comprehension. *Discourse Processes*, *26*, 159-190.

Sanford, A. J. (1985). Cognition and cognitive psychology. New York: Basic Books.

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

Tajfel, H. & Turner, J. C. (1979). An integrative theory of intergroup conflict. In M. A. Hogg & D. Abrams (Eds.) *Key readings in social psychology. Intergroup relations: Essential readings* (pp. 94-109). New York: Psychology Press.

Van Berkum, J. J., Brown, M., & Hagoort, P. (1999). Early referential context effects in sentence processing: Evidence from event-related brain potentials. *Journal of Memory and Language, 41*, 147-182.

Van Dijk, T. A. & Kintsch, W. (1983). *Strategies of Discourse Comprehension*. New York: Academic Press.

Vasishth, S., Von der Malsburg, T., & Engelmann, F. (2013). What eye movements can tell us about sentence comprehension. *Wiley Interdisciplinary Reviews: Cognitive Science, 4*, 125-134.

Waltz, D. (1975). Understanding line drawings of scenes with shadows. In P. H. Winston (Ed.), *The psychology of computer vision*. New York: McGraw-Hill.

Weinreich, U. (1953). *Languages in Contact*. The Hague: Mouton. ISBN 90-279-2689-1.

West C. & Zimmerman, D. H. (1987). Doing gender. *Gender and Society*, *1*, 125-151.

Zwaan, R. A. & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, *123*, 162-185.

Appendix A

Consent Form

Please consider this information carefully before deciding whether to participate in this research.

Purpose of the research: To examine the effectiveness reading comprehension.

Time required: Participation will take an approximated time of 60 minutes.

Risks: Low risks are associated with participating in this study.

Benefits:
If you there are any questions or concerns, please contact Adriana Garcia at adrianagarcia@dusty.tamiu.edu.

Confidentiality: Your participation in this study will remain confidential and there will be no link between your responses and your identity.

Participation and withdrawal: Your participation in this study is completely voluntary, and you may withdraw at any time without penalty. You may withdraw by informing the researcher that you no longer wish to participate (no questions will be asked).

Contact: If you have questions about this research, please contact Elva A. Garcia, 956-282-2321, adrianagarcia@dusty.tamiu.edu. You may also contact the faculty member supervising this work: Roberto Heredia, Ph.D., 956-326-2620, Texas A&M International University.

Whom to contact about your rights in this research or for questions, concerns, suggestions, complaints that are not being addressed by the research team, or in case of research-related harm: IRB Chair, Dr. Jennifer Coronado, irb@tamiu.edu or 956-326-2673.

Agreement: The nature and purpose of this research have been sufficiently explained and I agree to participate in this study. I understand that I am free to withdraw at any time without incurring any penalty.

Signature: _____ Date: _____

Name (print): _____

Appendix B

Counter Balance Sheet

1	A		21	B		41	D		61	C		81	A		101	B		121	D		141	C		161	A	
2	B		22	D		42	C		62	A		82	B		102	D		122	C		142	A		162	B	
3	D		23	C		43	A		63	B		83	D		103	C		123	A		143	B		163	D	
4	C		24	A		44	B		64	D		84	C		104	A		124	B		144	D		164	C	
5	B		25	D		45	C		65	A		85	B		105	D		125	C		145	A		165	B	
6	D		26	C		46	A		66	B		86	D		106	C		126	A		146	B		166	D	
7	C		27	A		47	B		67	D		87	C		107	A		127	B		147	D		167	C	
8	A		28	B		48	D		68	C		88	A		108	B		128	D		148	C		168	A	
9	D		29	C		49	A		69	B		89	D		109	C		129	A		149	B		169	D	
10	C		30	A		50	B		70	D		90	C		110	A		130	B		150	D		170	C	
11	A		31	B		51	D		71	C		91	A		111	B		131	D		151	C		171	A	
12	B		32	D		52	C		72	A		92	B		112	D		132	C		152	A		172	B	
13	C		33	A		53	B		73	D		93	C		113	A		133	B		153	D		173	C	
14	A		34	B		54	D		74	C		94	A		114	B		134	D		154	C		174	A	
15	B		35	D		55	C		75	A		95	B		115	D		135	C		155	A		175	B	
16	D		36	C		56	A		76	B		96	D		116	C		136	A		156	B		176	D	
17	A		37	B		57	D		77	C		97	A		117	B		137	D		157	C		177	A	
18	B		38	D		58	C		78	A		98	B		118	D		138	C		158	A		178	B	
19	D		39	C		59	A		79	B		99	D		119	C		139	A		159	B		179	D	
20	C		40	A		60	B		80	D		100	C		120	A		140	B		160	D		180	C	

VITA

Name: Elva Adriana Garcia

Address: 2801 Price St., Laredo, TX, 78043

Email Address: adrianagarcia@dusty.tamiu.edu

Education: B.A., Texas A&M International University, 2016, Psychology

Major field of specialization: Psychology