Exploring Linguistic Relativity: The Effect of the French Grammatical Gender System on Bilingual Adults' Perception of Objects

Zhuohan Chen
University of Oxford
St Cross College

Word Count: 19,877

Thesis submitted in partial fulfilment of the requirements for the Degree of
Master of Science in Applied Linguistics and Second Language Acquisition

August 2022

Candidate Number: 1059337

Table of Contents

Acknowledgements

I want to thank the existence of Linguistic Relativity, the starting point of my journey to applied linguistics. As written in my personal statement in the application to ALSLA, for me, language is never just a communication tool but a way of perceiving this world. The interplay between language and thoughts, or between bil/multilingualism and perception, is my long-term research interest and passion; it was, is, and will always be.

I am in outstanding debt to my supervisor, Dr Faidra Faitaki. She has been such a helpful, kind, organised, professional, brilliant, charming, and nourishing mentor in my Oxford journey. She teaches well, reviews my work meticulously, gives incredible suggestions for the research project, and is always there to help me with my questions. She helped me a great deal in recruiting participants and supported me at every opportunity. Whether it was applying for an RA position, DPhil, or conferences, she was always incredibly supportive. More importantly, her positive feedback, comments, praises, comfort, smiles, and generous compliments helped me keep going whenever I was low in mood or encountered any academic challenges. I am so lucky to have her as my supervisor and sincerely hope to express my gratitude to Faidra, a super role model to follow and to work with.

I also want to thank all the ALSLA teachers and professors. Thanks for teaching me how to be critical and how to be an independent researcher. Thanks, Oxford— 'The' dream place to apply and study. I would also like to thank my parents and friends for always supporting me. Thanks go to even those numerous scenes of sunset in the old Bodleian library, the gorgeous view in Radcliffe Camera, the Italiamo Café on the broad street, the trees in the Dept. of Education, and other Oxford places where I gain comfort mentally and physically. Also, I thank the countless cups of latte, cappuccino, flat white, as well as Doritos, hot wings from Dominos, Luoshifen, mint chocolate ice cream, cider, and diet coke, because, without them, this dissertation could hardly be born.

List of Tables

List of Graphs and Figures

Abstract

A growing body of literature on linguistic relativity (the hypothesis that language influences thoughts) has been focusing on whether there is an effect of the grammatical gender system on perception (Bassetti & Nicoladis, 2016; Lambelet, 2016; Samuel et al., 2019). Evidence suggests that there is an effect of the gender system on perception of both monolinguals and bilinguals (Bassetti & Nicoladis, 2016). However, the findings are potentially limited due to certain methodological limitations (Samuel et al., 2019), for example, the lack of L2 proficiency testing (Bassetti, 2007). Thus, attempting to apply a more rigorous methodology, this study aims to assess the potential effect of the French grammatical gender system on French speakers' and learners' perceptions of object gender. The study involved 140 participants, divided into four groups (N = 35 per group): English monolinguals, French monolinguals, English-French (English-dominant) bilinguals, and French-English (French-dominant) bilinguals. An online experiment was distributed to participants, including a background information questionnaire, English and French vocabulary tests, and a voice distribution task. Quantitative data were analysed using multi-level modelling. Follow-up open-ended question data were coded and analysed using chi-squared tests. The results supported linguistic relativity: The French grammatical gender system affected the perception of French monolinguals and English-dominant bilinguals. The effect of French on French-dominant bilinguals was not reduced by the acquisition of English and seemed to be independent of L2 proficiency. Additional findings included the potential tendency of French to introduce gender biases (stereotypical association between objects and gender, resulting from the grammatical gender system). The present study adds supporting evidence to the relativity debate by attempting to apply rigour by utilising a pilot study, a robust sample size, pertinent control items, L2 proficiency testing, and advanced data analysis tools. Pedagogically, findings highlight the need to emphasise the discrepancies between students' potential preconceived perceptions and the rules of the L2 grammar when teaching an L2 with a different grammatical gender system to L1. The study also shed new light on whether gendered languages draw out potential gender stereotypes among bilinguals.

1. Introduction

The past few decades have seen a resurgence of interest in linguistic relativity, the hypothesis that language influences thoughts (Athanasopoulos & Bylund, 2020). There is much debate surrounding linguistic relativity, specifically to what extent people's thoughts can be biased by the language (Casasanto, 2008; Pinker, 1994). The current prevailing approach appears to investigate how language relates to thought or not (Samuel et al., 2019). Recently, a growing body of research exploring linguistic relativity has been focusing on the effect of the grammatical gender systems on perception (Bassetti & Nicoladis, 2016; Lambelet, 2016; Samuel et al., 2019; Sato & Athanasopoulos, 2018). Various languages have systematic gender systems in grammar. For example, French has two genders (masculine and feminine) to mark nouns, while German has three (masculine, feminine, and neuter). Occasionally, the grammatical gender aligns with the animate nouns' gender. For example, in French, a boy ('un garçon') is masculine, and a girl ('une fille') is feminine. However, the grammatical gender system also assigns a gender to inanimate nouns. The linkage between the assigned gender and its referent's properties is often arbitrary, rendering the gender system challenging to learn (Bassetti & Nicoladis, 2016; Shimanskaya & Slabakova, 2019). For example, the *table* ('la table') in French is feminine, but the object does not have a prominent association with femininity. Therefore, studies have examined whether speakers of the gendered language would think of the objects' gender in line with the grammatical gender system (e.g., Bender et al., 2011).

The results of these studies show that there could be a potential effect of a language's grammatical gender system on its monolingual speakers' object perception, as well as on the object perception of the second language (L2) learners with a genderless first language (L1) (e.g., Lambelet, 2016). For example, a French monolingual speaker might perceive the object *table* as feminine, that is, in line with the grammatical gender in French. Likewise, an English learner of French might also think of *table* as feminine because of having acquired French. The impact of the grammatical gender system on speakers of a gendered L1 seems to remain

unaffected by the acquisition of a genderless L2 like English (e.g., Bassetti & Nicoladis, 2016). Thus, a French learner of English would still think of *table* as feminine. However, learning other grammatical gender systems might change this impact. For example, a French learner of German might think of the object table differently because *table* is masculine in German.

Nevertheless, certain methodological problems exist in relevant studies, including the absence of L2 proficiency testing, the confounding of gender stereotypes associated with the task, and insufficient control of items. It is difficult to conclude whether the effect of the gender system on perception still exists when the bilinguals recruited are the ones who can truly use the L2 effortlessly. Moreover, it is inconclusive whether the L1 grammatical gender system still impacts perception after confounding variables like gender stereotypes of objects are adequately controlled within the research design. It is also uncertain whether a higher level of L2 proficiency would increase, decrease, or have no effects on the impact of the L1 grammatical gender system. These methodological limitations render the results of previous studies potentially limited in terms of strength of inference.

Subsequently, the current study aimed to examine the effect of the French grammatical gender system (both as an L1 and L2) on perception by applying a methodology with rigorous additions (for example, controlling confounding variables in the research design). This study adopted a quantitative approach to answer four research questions:

1) To what extent do English and French monolingual speakers perceive objects' gender differently?
2) To what extent do English-French (English-dominant) bilingual speakers differ from English monolingual speakers in perceiving objects after learning a gendered language?
3) To what extent do French-English (French-dominant) bilingual speakers differ from French monolingual speakers in perceiving objects after learning a genderless language?
4) To what extent do English-dominant bilingual speakers differ from French-dominant bilingual speakers in perceiving the gender of objects?

The first question aimed to assess if (and if so, how) the grammatical gender system of French as an L1 affects French speakers' perception of objects. The second and third questions aimed to explore if (and if so, how) learning an L2 that has or does not have a grammatical gender system affects bilinguals' perception. Lastly, an additionally considered fourth question compared the effect of acquiring a gendered and a genderless L2.

This study consisted of an experiment that included a standardised L2 vocabulary test administered in English and French. The current study also attempted to carefully select control items based on the results of a pilot study to eliminate confounding of objects' stereotypical gender tendencies. The data of 140 participants in four groups (English monolinguals, French monolinguals, English-dominant bilinguals, and French-dominant bilinguals) were analysed through multi-level modelling and *t*-tests under three-way interaction effects (Field, 2018). Results showed supportive evidence of linguistic relativity, potentially indicating significant pedagogical implications concerning the teaching of French.

2. Literature Review

This chapter firstly presents the definition of bilingualism and linguistic relativity, with relevant discussion of literature. Detailed critiques of studies exploring the effects of grammatical gender on object perception are then provided, followed by a rationale for the study, research questions and predictions.

2.1. What It Means to Be Bilingual

2.1.1. Definition

'Bilingualism' has various interpretations and should be defined clearly. Within this dissertation, bilinguals are defined as people who use two languages fluently and effortlessly (Bialystok, 2001; Bloomfield, 1933); similarly, monolinguals are defined as individuals who can only use one language fluently and effortlessly. According to this definition, people who know a few utterances in another language, second language (L2) learners without sufficient proficiency in their L2, and people with casual knowledge of their families' home languages (i.e., heritage or minority languages as opposed to the majority language used in the community; Montrul, 2010) are not considered bilinguals (Bialystok, 2001).

2.1.2. Types of Bilinguals

Bilinguals differ in terms of the ages at which their L2 acquisition began. Simultaneous bilinguals receive input in two languages from early childhood, while sequential bilinguals only receive L2 input later in life. There is a controversy regarding when the early acquisition of two languages qualifies as simultaneous or sequential. McLaughlin (1978) believed simultaneous bilinguals are individuals who receive exposure to two languages before three. However, other researchers maintained that simultaneous bilinguals are individuals who receive two languages' input from birth (e.g., Padilla & Lindholm-Leary, 1984). As many empirical studies did not precisely describe their bilingual participants' onset of exposure (Romaine, 1999), this dissertation adopts McLaughlin's (1978) less strict criterion and considers people who acquired their L2 before three as simultaneous bilinguals. Bilinguals also differ in terms of their

proficiency in both languages. Balanced bilinguals are assumed to have equal lexical and grammatical abilities in both languages, while unbalanced bilinguals may perform better in their L1 or L2 (Grosjean, 1982). However, it should be noted that balanced bilinguals are challenging to find. In general, bilinguals appear to lag behind monolinguals in terms of linguistic (lexical) skills in both of their languages, depending on the quality (input) and quantity (dominance) of exposure to each language at home and school (Bosch & Ramon-Casas, 2014; Pearson et al., 1995; Paradis et al., 2021; Sebastián-Gallés et al., 2005).

2.1.3. Factors Affecting L2 Learning

Dominance plays an essential role in bilingual proficiency (Flege et al., 2002; Paradis et al., 2021). A dominant language is defined as the language in which "a bilingual has attained an overall higher level of proficiency at a given age, and/or the language that s/he uses more frequently, and across a wider range of domains" (Treffers-Daller & Silva-Corvalán, 2016, p. 4). Dominance is not equal to exposure. Language exposure or language contact, while closely related to dominance, refers to the amount (both quantity and quality) of linguistic input to which one is exposed (Matusevych et al., 2017; Paradis et al., 2021). However, language dominance differs from language exposure because it encompasses extra-linguistic factors such as language proficiency, preference, and the frequency of language use. Both language dominance and language exposure have been found to affect bilinguals' language proficiency (Flege et al., 2002; Matusevych et al., 2017). Therefore, they are included as variables in this project. Bilinguals are divided into an English-dominant group and a French-dominant group, and L2 exposure is used as a covariate.

As mentioned previously, the starting age of learning L2 can indicate whether the respondents are simultaneous or sequential bilinguals, which is linked to the effect of age on L2 learning (Hartshorne et al., 2018). The early age of acquisition has been found to play a role in L2 education (Sabourin et al., 2014). Sequential bilinguals have been found to have a longer processing time when they experience cross-linguistic influence (Kupisch et al., 2021).

Metalinguistic awareness (the ability to consciously reflect on the form and use of languages, see: Bialystok, 2001) has also been found to influence bilinguals' learning of the grammatical gender patterns of words (Brooks & Kempe, 2012). Thus, these variables should be investigated by obtaining participants' background information.

2.2. Language and Thoughts: Linguistic Relativity

The interaction between language and thought is an area of interest in the literature on bilingual language development (e.g., Pavlenko, 2003). In applied linguistics, the interaction between language and thoughts is usually referred to as the Sapir-Whorf hypothesis.

2.2.1. Definition

The Sapir-Whorf hypothesis, or Whorf(ian) hypothesis (Sapir, 1956; Whorf, 1956), has two versions. The strong version, Linguistic Determinism, claims that language *determines* thoughts; the weaker version, Linguistic Relativity, suggests that language *influences* thoughts. The film *Arrival*, where the character begins to see the future by learning an extra-terrestrial language, is an example of linguistic determinism. In this case, language acts as a barrier: The border of one language is the end of cognition, and cognition can be 'expanded' via the acquisition of another language. However, language in linguistic relativity is seen as more of a lens through which cognition passes (Ottenheimer, 2013). For example, 'black' tea in English is 'red tea' in Chinese. Due to the language difference, speakers of English and Chinese might perceive the colour of the same type of tea differently (before the tea leaves enter the water or afterwards). Most linguists agree that linguistic determinism is untenable due to its extreme position (Ahearn, 2017). If language truly limits cognition, speakers of Chinese would not be able to perceive that the 'red' tea is black before the water is added, which is not the case. By contrast, linguistic relativity is more acceptable and has been supported by empirical evidence (Ahearn, 2017).

2.2.2. Evidence

Empirical studies have explored linguistic relativity in various languages and across different domains. For example, cognitive linguistic research has focused on language and metaphor (Lakoff, 1980). Research on the intersection of psychology and the mind has concentrated on emotion, memory, motion event perception, and so forth (Flecken et al., 2015; Pavlenko, 2003; Perlovsky, 2009). The most frequently cited evidence for linguistic relativity could be on colour terminology (e.g., Athanasopoulos, 2009). For example, Brown and Lenneberg (1954) reported that Zuñi speakers had greater difficulties in distinguishing nuances in colours between blue and green than did English speakers. These difficulties were attributed to the lack of classification for blue and green colours in Zuñi. Similarly, concerning the effect of speakers' native languages, Chinese speakers tended to classify nouns by shape because of the prevalence of shape classifiers in Mandarin Chinese (Kuo & Sera, 2009). Likewise, Swedish speakers tended to estimate time based on length because Swedish describes time by distance terms such as 'long/short', whereas Spanish speakers tended to estimate time by size because Spanish describes time using quantity terms such as 'little/a lot' (Pandey, 2017).

Furthermore, linguistic relativity also applies to foreign language learning. A foreign language might affect the decision-making and reasoning processes because of the increased emotional and cognitive distance between the foreign language and the L1 (Chen, 2020; Keysar, 2012). For example, reading in a foreign language might decrease one's belief in conspiracy theories because that the greater emotional distance might lead to more rational thinking (Chen, 2020). As the cognitive function (decision-making and reasoning in this case) was influenced by adding another language, this line of research (foreign language effect) can also be seen as a form of linguistic relativity. However, it can be argued that these findings could be explained by factors other than the language; for example, psychological distance. Hence, foreign language effect research is often situated within the psychology domain instead of linguistic relativity. This project follows previous works exploring whether the French language affects

speakers' perception, both as L1 and L2. Therefore, linguistic relativity is the theoretical framework for this dissertation.

2.2.3. Criticism

Linguistic relativity has received much criticism (e.g., Bickel, 2000). Linguistic universalists (linguists who believe that all languages have the same underlying structure) have criticised the Whorfian hypothesis for being radical and have claimed that thoughts are independent of language (Pinker, 1994). Pinker (1994) argued that human beings do not think in the languages in which people communicate and that languages are fundamentally meaningless to human thoughts. However, Pinker's (1994) arguments against linguistic relativity have been criticised as committing the logical fallacy of attacking the strawman by mixing two interrelated but independent questions: "Do we think in language?" and "Does language shape thoughts?" (Casasanto, 2008, p. 63).

Another argument against linguistic relativity stems from the sociocultural perspective (e.g., Bickel, 2000): While language, to some extent, shapes thoughts, it might also be the case that the socio-cultural environment shapes language (Bidaoui, 2017). This sociolinguistic view can also complement linguistic relativity in that socio-cultural environment, language, and thoughts can be mutually intertwined. Most linguists adopt standpoints that are not based on one of the two extremes (linguistic determinism and universalism) but lie somewhere between them. In other words, there is a consensus that language may affect thinking, but different linguists differ in the degree to which they believe it does (Samuel et al., 2019).

As scholars explore linguistic relativity, some confusion regarding their definitions of 'thoughts' has arisen. Such confusion in the definition of 'thoughts' could be problematic for the relativity debate. Many terms have been used in the literature to denote the concept of thoughts, including ways of thinking, cognition, worldviews, cognitive processes, and perception (e.g., Athanasopoulos, 2009; Casasanto, 2008; Flecken et al., 2015). This dissertation argues that

some of these constructs are interrelated but are by no means identical (Firestone & Scholl, 2016) and thus cannot be used interchangeably. The definitions of these terms are provided in Table 1 (Oxford Languages, n.d.).

Table 1

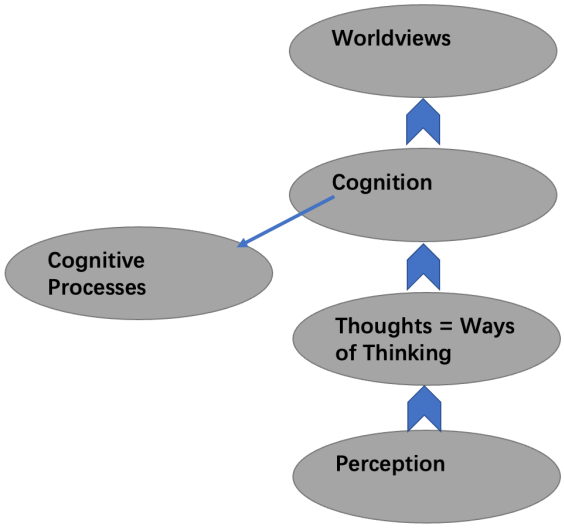
Definition of Different Versions of 'Thoughts

Term	*Definition*
Worldviews	"A particular philosophy of life or conception of the world"
Cognition	"The mental action or process of acquiring knowledge and understanding through thought, experience, and the senses"
Perception	"The way in which something is regarded, understood, or interpreted"
Ways of thinking	"The way in which your mind works, or the process of thinking about something"
Thoughts	"The action or process of thinking"
Cognitive processes	"A series of cognitive operations carried out in the creation and manipulation of mental representations of information"

Graph 1 shows the stratification of these definitions. *Worldviews* appear to be the highest stratum because it combines thoughts, perceptions, cognition, and philosophies of life. *Cognition* is the second, as it involves different thoughts about various things. According to definitions, *thoughts* and *ways of thinking* refer to the thinking process and can be considered interchangeable. *Cognitive processes* could be seen as a by-product of cognition: They are related to thoughts but might be a broader concept because many other brain activities are involved. *Perception* appears the closest to the objective world and can be the basis of *thoughts* which involves more subjective thinking. Therefore, in this dissertation, the *perception* was chosen as the key term to describe the potential conceptual changes resulting from the effect of language. The use of *perception* involves fewer controversies and might map better onto the construct of linguistic relativity on which the dissertation focuses.

Graph 1

Visualisation of Different Constructs



2.2.4. Bilingualism and Thoughts: A Bilingual Advantage?

The idea that language shapes thoughts, in turn, invites the appealing notion that acquiring more than one language affects cognition (Bassetti, 2007). Bilingualism has been claimed to improve cognitive abilities, such as executive function, working memory, and task-switching (Bialystok et al., 2014). However, these cognitive benefits have been questioned for two reasons. First, there is variation in the validity of studies testing the bilingual advantage, especially in the measurement of bilingualism and cognitive effects, and in the degree to which non-focal variables were controlled (Grosjean, 1989). For example, Antón et al.'s (2019) empirical study showed that the cognitive benefits detected in previous studies could have stemmed mainly from uncontrolled socio-demographic factors rather than bilingualism per se. The second reason is that a publication bias has been found to exist, favouring studies that support the bilingual advantage theory but disfavouring those that challenge it (de Bruin et al., 2015).

Regardless of the evidence in favour or against the bilingual advantage in cognitive abilities, there is a consensus that bilingual and monolingual development differ (Byers-Heinlein, 2014). Because of the notable differences in bilingual and monolingual development, investigating bilinguals' perceptions is intriguing regardless of whether cognitive differences exist. The effect of bilingualism on perception is investigated in this study; specifically, the project

explores the impact of acquiring a gendered or genderless L2 on the perception of objects. However, this project does not discuss whether the effect of the gender system on perception is advantageous or not because gender stereotypes might be introduced by the gender system (Bassetti & Nicoladis, 2016). It might then be difficult to decide whether such an effect is advantageous or disadvantageous.

2.3. The Effects of Grammatical Gender on Object Perception

2.3.1. The Grammatical Gender System

According to Bassetti (2007), cross-linguistic differences include three categories. The first encompasses aspects of reality (e.g., singular and plural; completed and uncompleted actions). The second contains categories along a continuum (e.g., colour terms). The grammatical gender system belongs to the third category, which is purely linguistic and independent of the real world. The grammatical gender system assigns genders to inanimate objects, thus providing a good case for linguistic relativity research. This case was considered because the gender system can be used to investigate the 'pure' linguistic effects on perception "without the potential confounding effects of non-linguistic cognition" (Bassetti, 2007, p. 254). The 'non-linguistic cognition' refers to the cognitive processes of the objective real-world. For example, people can perceive colours regardless of whether the languages they use have a unique terminology for a particular colour. The ability to perceive the colour does not take place on the linguistic level but will still affect how people attempt to describe the colour. Thus, colour terminology might not be useful evidence for linguistic relativity because the effect on colour perception does not stem purely from the language. However, this dissertation maintains that Bassetti's (2007) wording is overly strong, as potentially confounding effects exist regarding the socially stereotypical gender assigned to inanimate objects. For example, 'flower' is stereotypically related to the feminine gender. Nevertheless, in consonance with Bassetti (2007), the grammatical gender system is believed to provide a good testing ground for linguistic relativity.

2.3.2. Overall Effect

Many studies have supported the effect of the grammatical gender system on children's and adults' perceptions of object gender (Bassetti, 2007; Clarke et al., 1981; Flaherty, 2001). Nevertheless, the results are often tentative, as shown by a literature review conducted by Bassetti and Nicoladis (2016), in which the effect of the gender system on perception was found to rely on the pair of languages tested, the language proficiency of the participants, and the choice of tasks. Another systematic review conducted by Samuel et al. (2019) also claimed that the effect of the gender system on perception was context-and task-dependent. Samuel et al.'s (2019) review included 43 empirical studies on the effect of grammatical gender systems on perception from 1990 to 2018, with a total of 5,895 participants (monolinguals, bilinguals, and multilinguals). The review reported mixed findings: 38% and 28% of studies showed support or mixed support for the effect respectively, while the remaining 34% showed no support. Of all the tasks used in the included studies, voice attribution tasks (assigning a female or male voice to objects) and sex assignment tasks (assigning gender to objects) tended to yield the most supportive evidence. In comparison, the properties judgment task (describing an object's properties) yielded the most disapproving evidence against the idea that grammatical gender influences perception.

The review also concluded that the evidence (both supporting and disapproving) was subject to cross-linguistic, cross-cultural confounds, and other explanations (Samuel et al., 2019). For example, the strategies that the participants used when performing the task might be confounding. If they used the grammatical gender system in their L1 or L2 as a strategy when deciding on an object's gender, their practices would naturally lead to supporting evidence for the relativity hypothesis. In other words, when participants assigned gender to objects, they could notice the aim of the study and assign genders based on the grammatical gender of objects in a way to 'please' the researcher. That is, they may have been primed to provide the answer that they assumed the researcher wanted. A confound thus could emerge if the studies did not involve a means of controlling for this tendency. As Samuel et al.'s (2019) review included

both published and unpublished studies and mentioned clear eligibility criteria (empirical research involving human participants of real languages; reported in English) and synthesis processes, the overall findings are reliable.

2.3.3. Monolinguals' Perceptions Affected by a Gendered Language

According to linguistic relativity, speakers of a language with a grammatical gender system would consistently perceive the objects' gender with their language. Such an effect was found in Spanish (Konishi, 1993), Arabic (Clarke et al., 1981), French (Sera et al., 2002), Italian (Vigliocco et al., 2005), and Lithuanian (Vernich et al., 2017). However, the effect was less transparent in German, a language that has three grammatical genders (masculine, feminine, and neuter). Sera et al. (2002) found no effect of the German grammatical gender system on an object categorising task (a set of four objects with one having a different grammatical gender) performed by German children aged five, seven, and nine. In a later study, German adults were also found not categorising objects in line with the German gender system (Vigliocco et al., 2005). It is generally believed that the gender system in German tends to have a less potent effect because of the lexical-semantic complexity of its case system (Vigliocco et al., 2005; Zubin & Köpcke, 1984). Furthermore, using a neuter gender might have a weaker effect than languages that use a dichotomous gender system (Bassetti, 2007).

Flaherty (2001) conducted three consecutive experiments to explore how the grammatical gender system of Spanish affected the speakers' perception of objects. Flaherty's (2001) experiment investigated the effect of language on gender attribution, involving 144 native English-speaking and 144 native Spanish-speaking adults and children. The study used a robust design to examine the role of age: 48 adults, 48 children aged five to seven and 48 children aged eight to ten, with an equal gender distribution among participants. Participants were asked to assign a typical male or female name to black-white cartoons of 35 objects. Results revealed significant consistency between the grammatical gender in Spanish and the participants' choice of gender in the Spanish eight-ten-year-old children and adults, but not in five- to seven-year-

olds. Spanish children aged five to seven tended to assign gender to objects in line with their own gender or with famous characters in children's literature. By contrast, English participants marked the objects' gender according to their characteristics or properties. The results indicated that language affected the perception and that there was an age boundary (at around seven or eight), after which the language's effect began to play a role.

Flaherty's (2001) design was reliable because it allowed confounds to be eliminated before and after the primary test. An experiment was conducted before the aforementioned main study, focusing on the role of animacy in objects. Sixty-four native English adults completed a pre-experiment in which they assigned 20 objects a pronoun (she/he/it) in cloze tests, such as "The elephant is in the circus. __ is huge" (Flaherty, 2001, p. 20). The experiment revealed that inanimate objects received the pronoun *it* significantly more. Another experiment, conducted after the main study, focused on whether Spanish speakers rated objects based on objects' properties rather than grammatical gender. Thus, this experiment investigated the effect of objects' attributes on gender assignment. A semantic differential two-point scale of "low-high, hot-cold, small-big, beautiful- ugly, and sad-happy" (Flaherty, 2001, p. 25) was completed by 80 native English-speaking, and 80 Spanish-speaking adults and children (aged eight to ten); *high, cold, big, ugly,* and *happy* were deemed as masculine, whereas *low, hot, small, beautiful,* and *sad* were considered feminine. Results showed that the gender assignment and the attribute assignment were not significantly correlated for Spanish speakers, while the opposite was true for the English participants. Thus, the main finding that Spanish grammatical gender affected Spanish speakers older than seven appeared convincing.

Nevertheless, a potential limitation of Flaherty's (2001) study might be the gender stereotypes involved. In the third experiment, the semantic association between *sad/hot* and feminine properties or *happy/cold* and masculine traits were based on stereotypes. Such associations were arbitrary. It could be well argued that *happy/cold* is also associated with femininity. As a result of the arbitrary criteria for the semantic association task, the result of the third experiment

(Spanish speakers did not rate objects also based on objects' properties) was less reliable. Furthermore, the cumbersomeness of conducting three consecutive experiments could have been avoided if valid control items (objects that tend to be rated as 'masculine' or 'feminine' because of their properties) had been included in the main study.

2.3.4. Monolinguals versus Bilinguals: Learning a Gendered or Non-gendered Language

The studies presented in the previous section suggested that the grammatical gender system can affect the perceptions of monolinguals of a gendered L1. The effect is accentuated in the case of languages that only have a binary gender system. In line with research on bilingualism and thoughts, when the same object is assigned different genders in the two languages, bilingual speakers might have different concepts from monolingual speakers of each language (e.g., Bassetti, 2007; Bassetti & Nicoladis, 2016).

2.3.4.1. From Genderless to Gendered

Learning a grammatical gender system in a new language tends to affect the perception of native speakers of a genderless language (Bassetti & Nicoladis, 2016). For example, a longitudinal quantitative study by Kurinski and Sera (2011) showed that native English speakers' gender attribution to inanimate objects was affected after learning the Spanish grammatical gender system over one academic year. However, the effect for English learners of Spanish was less intense than it was for native Spanish speakers and did not increase in tandem with learners' proficiency in Spanish (Kurinski & Sera, 2011). This finding suggests that the acquisition of a gendered language might already affect learners' perception of a genderless L1 from the early stages of L2 acquisition, but the effect plateaus afterwards. A limitation of the study was that the native Spanish speakers who were recruited were Spanish-English bilinguals living in the U.S. Thus, the Spanish-English bilinguals' language dominance could have been affecting the effect – something that the authors did not explore or report. Despite this limitation, the study adopted a longitudinal design, which revealed that the effect did not go along with L2 proficiency individually (also see Wasserman & Weseley, 2009).

2.3.4.2. From Gendered to Genderless

Speakers of a gendered language also often learn an L2 that does not have a grammatical gender system. In this case, the effect of the L1 gender system on bilinguals with a gendered L1 and a genderless L2 appears to remain unaffected (Bassetti & Nicoladis, 2016). For example, in a study by Boroditsky and Schmidt (2000), German-English and Spanish-English bilinguals were found to perceive objects' gender based on their L1s, German and Spanish, despite their L2, English, being genderless. However, this study did not report on the English proficiency of these bilinguals. This made the interpretation of the results unclear, as the level of proficiency at which the effect plateaus and whether the effect changes because of having high L2 ability were unknown. The same 'unwavering' impact of learning a genderless L2 as a learner with a gendered L1 was also found in French-English bilingual adults and children, who aligned their gender perception of objects with the French grammatical gender system (Forbes et al., 2008; Nicoladis & Foursha-Stevenson, 2012; Sato et al., 2013).

Whether the effect of the gender system decreases or increases with L2 proficiency is inconclusive. Sato et al. (2013) reported that French-English bilingual adults' stereotypical gender attitudes were impacted less by the French grammatical gender system as their English proficiency increased. However, Kurinski and Sera's (2011) found that an increase in Spanish ability did not affect the perception of English learners of Spanish. Sato et al.'s (2013) findings might be more reliable than those of Kurinski and Sera (2011) because Sato et al. (2013) adopted a standardised c-test to establish L2 proficiency and divided their participants into two groups (advanced and intermediate) more objectively. By contrast, Kurinski and Sera (2011) did not test the learners' Spanish levels; they assumed that they were advanced because they were students in the Department of Spanish. The lack of proper L2 proficiency testing (as opposed to participants' self-evaluation) has been a common limitation in previous studies. This project explores proficiency's role in the acquisition of the grammatical gender system further.

2.3.4.3. Having More Than one 'Gender'

The question of whether the L1 would be affected by the L2 has also been the focus of research regarding having two gendered languages. Phillips and Boroditsky (2003) maintained that language dominance was crucial. In their study, bilinguals with two gendered languages (Spanish-German and German-Spanish bilinguals) rated objects with opposite genders in German and Spanish consistently with the gender in their dominant language. However, this study also suffered from the common limitation of participants self-reporting their 'more proficient language' without testing.

Bassetti (2007) and Lambelet (2016) showed that adding another gender system tends to weaken the effect of the L1 gender system. Bassetti (2007) investigated whether Italian-German bilinguals and Italian monolinguals had different concepts of the same object when they had opposite genders in German and Italian. Twenty-one Italian-German bilingual and 21 Italian monolingual nine-year-old children (control) participated in an online voice attribution task in which they assigned a male or female voice to 12 concrete objects (choosing from two voice files). All the children were native Italian speakers living in Italy. Results showed that grammatical gender significantly affected children's perception of objects, as indicated by monolingual children's preference to assign a gender consistently with the Italian grammatical gender system. The same result was found by Boutonnet et al. (2012), who also used event-related brain potentials. However, the choices of the Italian-German bilingual children and monolingual Italian children differed (Bassetti, 2007). This finding indicates that the addition of the German grammatical gender system might influence the effect of the original Italian one.

Bassetti's study (2007) was good in carefully selecting items, with 12 objects having no evident association with male or female and two control objects that have. Nevertheless, this merit cannot compensate for the often-appeared limitation: The author assumed the children were bilinguals without any proficiency testing. The children were considered bilingual if they self-

rated as having no preference for speaking Italian or German, stated that they were equally good at both languages, and had native-like speaking proficiency in both. The self-deception bias in self-reported data could not be ignored (Dörnyei & Taguchi, 2009). Whether the children fully understood what it meant to be 'bilingual' or have 'equal/native-like proficiency' was questionable. The study's results can thus only be accepted tentatively because the recruited 21 Italian-German bilingual children might not be technically bilinguals.

Similarly, Lambelet (2016) discussed the difficulties bilinguals encounter when learning a second grammatical gender system. A voice attribution task was distributed to 282 adults French L2 learners with 21 different L1 backgrounds including languages with binary or tertiary grammatical gender systems (masculine and feminine; masculine, feminine, and neutral) and languages without it. Participants chose a female or male voice for 10 inanimate objects, followed by a good distractor to decide whether the voice belonged to an old or young person. A significant effect of the L1 grammatical gender on the participants' voice attribution task performance was found. Moreover, learning an L2 grammatical gender was found to weaken the impact of L1 grammatical gender: Participants who had a higher proficiency in the French grammatical gender system were less likely to assign voices following the object's grammatical gender in their L1.

The study included a second task, French grammatical gender identification, in which participants were required to identify the French gender of objects used in the experiment. The task found that the conjectured French grammatical gender played a role in deciding gender. For example, *beer* is neutral in German ('das Bier'), but, if a German participant incorrectly thought *beer* in French was masculine (in fact, feminine), s/he would assign a feminine gender, indicating the effect of conjectured L2 grammatical gender system. The findings should be interpreted cautiously because no French proficiency tests were implemented to assess the participants' French level. A simple vocabulary decision task could have made a difference compared to the self-evaluation of proficiency. Also, objects did not involve reasonable control

of items with a stereotypical gender, which jeopardises the interpretation of the effect of the L2 gender system. Furthermore, choosing a masculine or feminine voice may be too dichotomous. It might be difficult for participants to decide on the gender of an inanimate object as entirely masculine or feminine. A better method could be to give participants more space to choose, as they might tend to decide on somewhere in between (but not entirely neutral).

2.4. Research Gap and Rationale for the Study

In summary, the grammatical gender system appears to affect perception, but its effect depends on contexts and tasks. A binary grammatical gender system appears to affect the perception of monolingual speakers and learners with a genderless L1. However, the L1 tends to remain unaffected by learning a genderless L2 but might be weakened by acquiring another gendered L2. The results of relevant studies are often inconclusive and tentative due to methodological problems, including the lack of L2 proficiency testing, the use of gender stereotypes, and the lack of control items for objects' stereotypical gender tendencies. Therefore, this project aims to explore the effect of the French grammatical gender system with improved methods.

The project can be important for multiple reasons. Firstly, it explores the weaker Sapir-Whorf hypothesis, linguistic relativity, which has been the subject of intense debate over the years. Secondly, the study considers the aforementioned methodological criticisms and implements an improved design. For example, the respondents' L2 proficiency was measured to compensate for the limitations noted in previous literature. Moreover, the project carefully selects control items from a pilot study to eliminate the confounding gender stereotypes associated with objects on the item list. Lastly, the study can provide relevant pedagogical implications regarding the teaching and learning of French gender pronouns.

2.5. Research Questions

The aims of this project are threefold. Firstly, it investigates if (and if so, how) the grammatical gender system in French affects French speakers' perception of objects (e.g., is the *table*

perceived as masculine or feminine?) compared to English monolingual speakers. Secondly, it explores if (and if so, how) learning an L2 that has/does not have the grammatical gender system affects bilingual adults' perception of objects compared with monolingual speakers. Thirdly, it tests if the patterns of object perception between English-dominant bilingual adults and French-dominant bilingual adults are different and, if so, how.

The research questions (RQ) are as follows:

1. To what extent do English and French monolingual speakers perceive objects differently (assigning voice; masculine or feminine)?
2. To what extent do English-French bilingual speakers differ from English monolingual speakers in perceiving objects after learning a gendered language?
3. To what extent do French-English bilingual speakers differ from French monolingual speakers in perceiving objects after learning a genderless language?
4. To what extent do English-dominant bilingual speakers differ from French-dominant bilingual speakers in perceiving the gender of objects?

2.6. Predictions

The study involved two within-participants variables, the chosen objects' French gender (male or female) and condition (experimental items and control items), and one between-participants variable (language group). For RQ1 and RQ2, a main effect of group was expected; English and French monolingual speakers, as well as English monolinguals and English-dominant bilinguals, were expected to differ in their ratings of objects' gender. An interaction effect between group and condition for both RQ1 and RQ2 was also expected: French monolinguals and English-dominant bilinguals would rate control items (measured more stereotypically-gendered concepts) based on their characteristics or properties and experimental items based on the grammatical gender system in French. By contrast, the English monolingual group was predicted to rate all items based on their characteristics or properties. Moreover, an interaction effect between group and gender was expected, suggesting that French monolinguals and

English-dominant bilinguals would rate male objects lower and female objects higher than English monolinguals.

No main effect of the group was expected for RQ3. As shown in previous literature, learning a genderless language did not appear to influence the impact of the L1 grammatical gender system (e.g., Bassetti & Nicoladis, 2016). Thus, the French-English bilingual group was not expected to differ from the French monolingual group. No interaction effect between group and condition, nor between group and gender was predicted. For RQ4, no predictions were made because previous work suggests that L2 proficiency might play a role (e.g., Sato et al., 2013). Since participants' L2 proficiency might be various and should be treated as a covariate, the prediction of whether the two bilingual groups differ is thus hard to make.

3. Methodology

The study adopted a quantitative approach with an experimental design (Rogers & Révész, 2020), aiming to explore any relationship between the French grammatical gender system and participants' perception of objects' gender. The entire experiment consisted of three parts: background information, the L2 proficiency measure, and the objects' perception measure. During the experiment, the participants were occasionally asked to explain their selection in the voice attribution task.

3.1. Participants

Given the RQs, the study required four groups of adult participants: English monolinguals, French monolinguals, English-French bilinguals (English-dominant), and French-English bilinguals (French-dominant). Determining the number of participants was based on the power the study hoped to achieve. According to Cohen (1977), the threshold of power should be set by considering the relative risk of rejecting the null hypothesis incorrectly (type I error, α) compared to falsely accepting the null hypothesis (type II error, β) (Bababekov et al., 2019). Since α is commonly set at 0.05, β would be 0.2, resulting in (1-β)=0.8 as a threshold for power. According to a power analysis using the G*Power 3 programme (Faul et al., 2007), a total of 134 participants (33 per group) would be required to achieve 80% power at the α=.05 level [(1-β)=.8] for main effects[1] (see Appendix H).

A total of 140 participants (35 per group) were included in the study. They were categorised into four groups based on their linguistic backgrounds. Initially, 214 participants completed the experiment within 45 minutes, including 36 English monolingual speakers, 34 French monolingual speakers, 46 English-dominant bilingual speakers, and 101 French-dominant bilingual speakers. French-dominant bilingual speakers outnumbered the other three groups because they were mainly found in the pool of French-dominant bilinguals[2]. After the manual

[1] Note that an effect size of .29 was used based on Bassetti (2007), discussed in detail in the previous chapter.
[2] French monolinguals were difficult to find. However, they could be retrieved from the pool of French speakers who regarded themselves as French-English bilinguals but in fact, only had a low level of English proficiency (defined as significantly lower than B1).

screening, two participants who declared themselves French-English bilingual speakers were categorised as French monolingual speakers because their tested English proficiencies were too low (around A1/A2 level). Similarly, two participants who declared themselves as English-dominant bilinguals were moved into the English-monolingual group due to their low levels of French. Four participants' data were excluded because they failed the attention checks (details provided in section 3.2). The data for eight French speakers having an L2 other than English and two English speakers having an L2 other than French were deleted to ensure the sample's homogeneity and the data's validity.

Valid data included 37 English monolingual speakers, 35 French monolingual speakers, 45 English-dominant bilinguals and 97 French-dominant bilinguals. To avoid the effect of unequal sample size on data analysis (Shaw & Mitchell-Olds, 1993), the data for 35 participants were selected randomly from the remaining three groups via Cluster Random Sampling (Frey, 2018), assuming the language differences were homogenous characteristics that differentiated groups from each other (forming 'clusters'). The random selection was realised via wheelofnames.com, a free online random selector. With 140 participants, the study can be deemed as having sufficient power according to the aforementioned power analysis. The profile of the final participants is presented in Table 2.

Table 2

Participants' Profile

Group	***N***	***Gender***		***Age***			***Education Level***			
		M	***F***	***18-30***	***31-50***	***>50***	***Middle***	***Tertiary***	***Vocational***	***Unspecified***
English Mono.	35	16	19	13	18	4	6	19	9	1
French Mono.	35	20	15	24	9	2	3	28	3	1
Eng-dom. Bil.	35	16	19	14	13	8	3	25	7	/
Fre-dom. Bil.	35	17	18	21	12	2	3	31	1	/
Total	140	49%	51%	51%	37%	11%	11%	74%	14%	1%

As shown in Table 2, the participants' gender distribution was relatively equal, thus mitigating the effect of participants' own gender on their perception (Flaherty, 2001). Regarding bilingual participants' L2 information, 10 (14%) were simultaneous bilinguals (began learning L2 before three years old), while the remainder, 39% and 47%, began learning the L2 at the ages of four to nine and above ten. Regarding metalinguistic awareness, 58% of bilinguals responded that they had had other language learning experiences, mainly German and Spanish. Regarding L2 exposure, most bilinguals had 0-30% daily exposure to the L2 (59%), with the remaining 27% and 14% of bilinguals having 30-70% and more than 70% of exposure, respectively. Moreover, 36% of the bilinguals had L2 qualifications, such as a BA in French Literature or working status for the government. However, as shown in Table 3, there was a discrepancy between English-dominant bilinguals' French level and French-dominant bilinguals' English level: Most of the French participants were advanced/proficient in English (C1 and C2), while most of the English participants only had a low intermediate level (B1). The dominance of English worldwide might explain this discrepancy (Huang, 2016).

Table 3

Bilinguals' Proficiency in the L2

Group	*Tested Proficiency in L2*		
	Advanced /Proficient (C1&C2)	*Upper Intermediate (B2)*	*Lower Intermediate (B1)*
English-dominant bilinguals' French	3	15	17
French-dominant bilinguals' English	24	10	1

3.2. Materials

3.2.1. Piloting

The entire experiment was piloted with 30 English and 30 French monolingual speakers who were not included in the subsequent experiment. Table 4 shows the mean ratings (with S.D.) of English and French monolinguals for each object, the condition and French gender, and objects

rated significantly differently by participants (marked with **). The piloting aimed to ensure internal validity: The rating participants needed to give for objects were based on personal evaluation, not on the stereotypical association with the object's characteristics or properties (for example, *necklace* is always attached to feminine, while *hammer* is usually masculine, Bassetti, 2007). Thus, piloting allowed the study to determine the objects with an inherent 'gender' attribution (which would correspond to control items in the experiment) and those without (experimental items).

Table 4

Piloting Results

French Gender	*Condition*	*Objects*	*Participants*	*M*	*S.D.*	*French Gender*	*Condition*	*Objects*	*Participants*	*M*	*S.D.*
Feminine (n=16)	Experimental (n=11)	Baguette	English	43.73	25.0	Masculine (n=16)	Experimental (n=11)	Bed	English	55.3	18.3
			French	34.93	17.8				French	46.8	21.2
			Total	39.33	21.9				Total	51.1	20.1
		Banana	English	47.27	18.3			Fish	English	54.9	16.7
			French	47.07	23.4				French	45.4	20.8
			Total	47.17	20.8				Total	50.2	19.3
		Cliff	English	31.3	21.1			Milk	English	49.8	18.9
			French	41.2	24.4				French	42.8	14.3
			Total	36.25	23.2				Total	46.3	17.0
		Cup	English	58.77	16.5			Watch	English	44.8	23.1
			French	50.9	20.5				French	47.9	25.8
			Total	54.83	18.9				Total	46.4	24.3
		Fork	English	39.57	18.6			Champagne**	English	75.8	15.0
			French	49.87	19.9				French	49.8	21.3
			Total	44.72	19.8				Total	62.8	22.5
		Frog	English	34.93	18.6			Cheese**	English	57.2	16.8
			French	42.07	19.0				French	41.8	17.8
			Total	38.5	19.0				Total	49.5	18.9
		House	English	46.17	20.1			Cloud**	English	52.4	19.1
			French	50.1	22.6				French	39.2	14.8
			Total	48.13	21.3				Total	45.8	18.2
		Moon	English	55.8	26.7			Elephant**	English	43.9	22.4
			French	63.37	25.5				French	27.8	18.3
			Total	59.58	26.2				Total	35.9	21.8
		Church**	English	37.03	22.6			Fruit**	English	65.3	15.2
			French	53.27	20.8				French	52.3	20.2
			Total	45.15	23.0				Total	58.8	18.9
		Mountain**	English	31.53	21.4			Tree**	English	60.3	19.3
			French	43.97	22.3				French	45.1	16.5
			Total	37.75	22.5				Total	52.7	19.4
		Table**	English	36.57	20.9			Sun**	English	60.9	23.0
			French	48.67	24.3				French	45.7	21.2
			Total	42.62	23.3				Total	53.3	23.2
	Control (n=5)	Candle	English	60.7	16.7		Control (n=5)	Airplane	English	38.5	21.4
			French	61.1	17.5				French	26.6	16.0
			Total	60.9	17.0				Total	32.6	19.7
		Car	English	33.9	20.2			Bike	English	38.1	17.5
			French	33.8	22.5				French	34.4	13.0
			Total	33.85	21.2				Total	36.3	15.4
		Teapot	English	68.93	17.4			Bird	English	68.7	17.1
			French	65.43	20.6				French	73.1	15.9
			Total	67.18	19.0				Total	70.9	16.5
		Mouse	English	72.83	16.8			Snake	English	36.1	22.8
			French	61.37	20.1				French	34.0	21.8
			Total	67.1	19.3				Total	35.1	22.1
		Tower	English	30.63	16.5			Train	English	29.9	15.2
			French	36.97	16.5				French	35.4	19.1
			Total	33.8	16.7				Total	32.7	17.4
			English	48.63	25.0				English	51.0	22.6
	Feminine Total		French	51.71	24.0		Masculine Total		French	42.8	21.1
			Total	50.17	24.6				Total	46.9	22.3
									English	49.8	23.9
				Total					French	47.2	23.1
									Total	48.5	23.5

***objects rated significantly differently*

3.2.2. Objects.

Based on the pilot, the final objects (see Appendix F for object list) included in the experiment were 32 common concrete objects (16 feminine and 16 masculine words in French), some of which were based on Flaherty's (2001) work. All words had high frequency (*M*=6791) in British National Corpus, and the mean length was 1.38. Ten control items were chosen (English and French monolinguals rated them <= 40 or >= 60), with five in each gender. The remainder, 11 feminine and 11 masculine objects, were experimental items. The ten experimental items rated significantly differently by participants during the pilot were chosen to be the target of open-ended questions (in which participants explained their choices in a text box).

3.2.3. Background Information

Participants first completed a background information questionnaire (see Appendix B). Even though participants might be unable to give accurate self-descriptions, questionnaires are the most practical and direct way of obtaining background information from participants (Dörnyei & Taguchi, 2009). The background information collected pertained to the participants' language backgrounds, language exposure, age of onset for L2, self-perceived L2 proficiency, language qualifications, gender, education level, and other language learning experiences likely to be moderating variables in this study.

3.2.4. Vocabulary Tests

The second part of the experiment included vocabulary tests measuring the bilinguals' proficiency in English or French. English-dominant bilinguals completed the French vocabulary test in English, while French-dominant bilinguals completed the English tests in French. Vocabulary was used to measure proficiency because receptive vocabulary knowledge is a good predictor of L2 proficiency (Miralpeix & Muñoz, 2018). The vocabulary test also assisted in the manual screening of bilingual participants: A participant needed to attain a vocabulary test score of at least 60% (B1) to be deemed bilingual instead of monolingual.

The vocabulary test used for English was the Lexical Test for Advanced Learners of English (LexTALE). LexTALE is a visual lexical decision task (for the list of words, see Appendix D), proven to be a good predictor of one's vocabulary knowledge or even general proficiency in English (Lemhöfer & Broersma, 2012). Brysbaert (2013) developed a French version of LexTALE (LexTALE_Fr) in parallel to the English version (see Appendix E). On average, LexTALE only takes 3.5-5 minutes to complete. Despite its brevity, it correlates well with other valid proficiency tests (Lemhöfer & Broersma, 2012). In both versions, two-thirds of the words are actual words, while the remaining one-third are non-words. The English LexTALE contains 60 trials altogether (40 words and 20 non-words), while the French LexTALE consists of 84 trials, with 56 words and 28 non-words. French LexTALE had its size expanded in case some items needed to be dropped. Actual words in English LexTALE had a frequency of between 1 to 26 per million (pm) words. In the French version, 30% of actual words had a frequency of < 1 pm, 49% between 1-10 pm, and 27% between 10-100 pm. The French LexTALE asks test-takers to choose the words they know (Brysbaert, 2013), while the English LexTALE asks test-takers whether the word they see is real or not (that is, involves a yes/no decision). This study adapted the French version to include the same yes/no decision task as the English version to ensure that the administration of the two measures would be the same.

English LexTALE was only given to French-dominant bilingual speakers and vice versa. Words were presented individually, and the order was randomised for each participant. The scoring of LexTALE includes "the correct percentage measure but corrected for the unequal proportion of words and non-words by averaging the percentages correct for these two item types" (Lemhöfer & Broersma, 2012, p. 329). The formula is: *((number of words correct/number of real words*100) + (number of nonwords correct/number of non-words*100)) / 2.* However, the platform on which the experiment was administered could not use the scoring formula but the pure, correct rate. Thus, all 70 bilinguals' L2 proficiency was scored manually. The relationship between LexTALE scores and L2 proficiency in this study was based on Lemhöfer and Broersma's (2012) work, as shown in Table 5.

Table 5

Relationship between LexTALE Scores and L2 Proficiency

CEF Level	*Description*	*LexTALE Score*
C1 & C2	Upper & lower advanced/ proficient user	80%-100%
B2	Upper intermediate	60%-80%
B1 and lower	Lower intermediate and lower	Below 59%

Table adapted from Lemhöfer and Broersma (2012)

3.2.5. Voice Decision Task

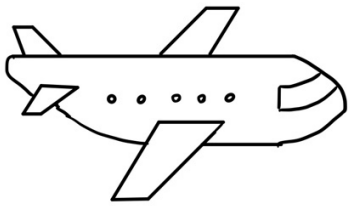
In the third part of the experiment, participants completed a voice decision task in which they decided on the voice for objects (Lambelet, 2016). For example, 'if the *airplane* could speak, what kind of voice would it have?'. In the French version, instead of *le avion* ('the airplane'), *l'object* ('the object') was used to avoid the hint of the gender pronoun. As shown in Figure 1, participants attributed voices using a slider, with the far-left representing low-pitch (or 'masculine' sound) and the far-right representing high-pitch (or 'feminine' sound). The slider gave rise to continuous data, as the far-left corresponded to 1% and the far-right to 100%. Thus, the point where the participants placed the slider represented a percentage. An object was considered to have a masculine voice if a participant assigned it a score of up to 50%, and the object was considered to have a feminine voice if the participant gave it a score of over 51%. The slider also allowed more space for participants to choose from dichotomous masculine or feminine, thus potentially making them feel more comfortable.

All objects were illustrated using black-white simple-line drawings drawn by the researcher, controlling for complexity (see Appendix O). The black and white format was chosen to eliminate the gender connotations of colour (Flaherty, 2001). Two ALSLA students checked their recognisability, resulting in changes to the drawings for one object, *table*. Unlike other studies using voice attribution tasks, no audio files were used in this experiment. This decision was motivated by piloting with audio files showing participants were disturbed by the voice in

the file. Moreover, as the task included numerous objects, participants might feel fatigued from clicking two voice files following each object.

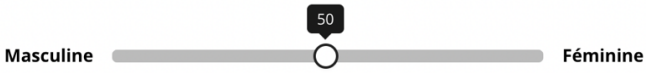
Figure 1

An Example of Voice Attribution Task



If the **Airplane** could speak, what kind of voice would it have? (English version)

Si **l'objet** pouvait parler, quel genre de voix aurait-il ? (French version)



Ten open-ended questions (serving as attention checks) were presented in the voice attribution task following the ten objects (rated significantly differently in the piloting). The question asked, 'why do you think the voice you just rated will be like this?'. Participants had to type their answers in a box.

3.3. Procedure

3.3.1. Translation Fidelity

The experiment had two versions, one in English and one in French. English monolinguals and English-dominant bilinguals completed the experiment in English, while French monolinguals and French-dominant bilinguals did so in French. Thompson and Dooley (2019) suggested back-translation to ensure the highest translation fidelity. Thus, the materials were translated by two independent forward- and backward translator tools (Youdao Translator and Google Translator). The back translations were then compared to the original texts. The final translation was checked by an expert in French (a French C2 learner who had obtained a BA and MA in French), resulting in a few changes to the wording.

3.3.2. Obtaining Consents

At the beginning of the experiment, the English-French bilingual information sheet and the consent form (see Appendix A) were presented to the participants. Participants had to declare their willingness to participate and tick the box 'I'm 18 years or above' to continue. The consent form explained how they could withdraw (by closing the browser), how the data was non-identifiable, and how they could contact the researcher if they had questions or concerns.

3.3.3. Administration

The experiment link was compiled and generated on the experiment builder Gorilla. The primary source for participants was Profic.com, a participant recruitment website using pre-screening criteria. For example, if French-dominant bilingual participants are needed, pre-screener conditions can be: 'having French as L1', 'fluent in two languages/bilingual', and 'fluent in English'. Snowball sampling was also adopted: Information about the study was disseminated through mailing lists, friend contacts, university societies, social media, and the French Institute in London, among other sources. Potential participants were provided with two research flyers (see Appendix N, one in English and one in French). Participants recruited via snowball sampling were incentivised by the chance of winning a £50 Amazon voucher.

Before the piloting, the study link was tested several times to ensure that it works on different platforms (computers, tablets, phones). The project supervisor then reviewed the link and provided suggestions regarding the jump logic. Following this, the link was sent to a group of the researcher's contacts outside the field to ensure the phrasing of the questions was understandable. This resulted in substituting the term 'foreign language/second language with the phrase 'the language in which you are most fluent apart from L1' in the questionnaire. Then, the experiment was piloted with 60 participants (the same pilot discussed earlier), resulting in redrawing two objects, omitting four objects, and substituting two objects. These changes were made to ensure the recognisability of objects and an even number of experimental and control

items in each French gender. The experiment was then implemented following a final, personal check.

3.3.4. Ethical Considerations

Data collection and protection followed the University's Guidance on Data Protection and Research (University of Oxford Central University Research Ethics Committee, 2018). The participants were recruited to complete the online experiment after receiving ethical approval for the experiment (see Appendix G). As essential ethics practices include anonymisation, consent acquisition, and information disclosure (Tao et al., 2017), participants were provided with consent and information forms, and all data collected were non-identifiable through the integrated encryption system of both Gorilla and Prolific.com that does not allow for private information about the participants to be disclosed. As recommended in the CUREC Best Practice Guidance 06 for Internet-mediated Research (University of Oxford Central University Research Ethics Committee, 2021), data, including consent records and processed (coded and statistically analysed) data, were stored electronically in a password-protected file in a portable hard-drive.

This project might be sensitive due to simply categorising voice as having low-pitch vowel sound (generally perceived as the quality of masculine) and high-pitch vowel sound (usually perceived as the quality of feminine). To reduce the risk of disagreement from participants, the consent form stated explicitly that 'this study generally categorises masculine and feminine voices in a continuum for purely academic purposes. Also, a 'prefer not to say' option was given for each background information question.

4. Results

This chapter first presents the overall main and interaction effects of the three independent variables, followed by the results for each research question (RQ). The outcome measure was the 140 participants' ratings for 32 objects. There were predictors regarding the objects were Condition (experimental and control, coded as 1 and 0) and ObjectGender (French female or male, coded as 1 and 0). There was also one predictor pertaining to the individuals LanguageGroup (the English monolingual group was coded 0, French monolingual group was coded 1, English-dominant bilingual group was coded 2, and French-dominant bilingual, 3). The group, grammatical gender and condition were categorical variables, while the dependent variable was continuous.

Each participant's ten open-ended questions (1400 answers) were coded according to three levels: 'grammatical gender', 'characteristics and feelings', and 'other'. For example, if a participant's voice rating for *table* was 'le mot en français est féminin' ('the word is feminine in French'), it was coded as a 'grammatical gender' answer; if the answer was 'table is masculine as it is a sturdy object', it was considered to be a 'feelings' answer and, if the answer was 'unsure/I don't know', it was classed as an 'other' answer. A second coder within the Department of Education at the University of Oxford rated all the answers again to reduce the researcher bias. The inter-coder reliability (the consistency in the coding between two coders) was indicated by Cronbach's $\alpha = 0.87 > 0.8$ (Appendix I), which means that the coding results were not affected by researcher bias (Field, 2018).

4.1 Overall Effects of Independent Variables

Outliers in the responses for each group were identified via scatterplots (see Appendix K). Because outliers are not representative sample data that can skew the distribution, they were reassigned to the group's next highest score (not an outlier); that is, they were winsorised (Field, 2018). After reassigning the outliers, the distributions of each group for each variable were found to be normal, as the absolute values of skewness and kurtosis were between -2 and +2

(George & Mallery, 2010) as shown in Table 6.

Table 6

Descriptive Statistics

Group		*N*	*Min*	*Max*	*M*	*S.E.*	*S.D.*	*Skewness*		*Kurtosis*	
								Statistic	*Std. Error*	*Statistic*	*Std. Error*
English monolingual	Ratings for masculine objects	35	507.0	855.0	716.3	12.9	76.5	-0.4	0.4	0.3	0.8
	Ratings for feminine objects	35	683.0	954.0	838.7	11.2	66.5	-0.1	0.4	0.1	0.8
	Ratings for experimental objects	35	860.0	1264.0	1086.8	17.3	102.0	-0.4	0.4	0.1	0.8
	Ratings for control objects	35	351.0	548.0	458.1	9.2	54.6	-0.1	0.4	-0.4	0.8
	Ratings for all objects	35	26976.0	671667.0	385639.3	29292.5	173296.8	0.1	0.4	-0.4	0.8
French monolinguals	Ratings for masculine objects	35	634.0	1268.0	933.9	27.6	163.0	0.5	0.4	-0.5	0.8
	Ratings for feminine objects	35	393.0	909.0	704.6	20.7	122.3	-0.5	0.4	0.0	0.8
	Ratings for experimental objects	35	849.0	1379.0	1125.1	23.3	137.8	0.0	0.4	-0.4	0.8
	Ratings for control objects	35	428.0	614.0	523.3	10.2	60.5	0.1	0.4	-1.1	0.8
	Ratings for all objects	35	15988.0	716105.0	363141.3	30097.2	178057.7	-0.1	0.4	-0.3	0.8
L1 English L2 French	Ratings for masculine objects	35	555.0	1316.0	884.8	38.1	225.7	0.6	0.4	-0.7	0.8
	Ratings for feminine objects	35	367.0	938.0	716.1	27.4	161.9	-0.9	0.4	0.0	0.8
	Ratings for experimental objects	35	949.0	1252.0	1112.9	15.6	92.3	0.1	0.4	-0.9	0.8
	Ratings for control objects	35	324.0	607.0	485.4	12.2	72.3	-0.3	0.4	0.0	0.8
	Ratings for all objects	35	12302.0	803857.0	329500.2	39598.6	234268.3	0.8	0.4	-0.3	0.8
L1 French L2 English	Ratings for masculine objects	35	643.0	1250.0	897.2	27.0	159.6	0.8	0.4	0.1	0.8
	Ratings for feminine objects	35	452.0	846.0	711.9	18.0	106.3	-0.8	0.4	0.3	0.8
	Ratings for experimental objects	35	872.0	1309.0	1104.6	18.3	108.4	0.1	0.4	-0.4	0.8
	Ratings for control objects	35	393.0	643.0	508.4	11.4	67.3	0.3	0.4	-0.3	0.8
	Ratings for all objects	35	22261.0	989887.0	314549.1	44305.9	262117.4	0.9	0.4	0.0	0.8

Normal distribution enables parametric tests. Thus, conducting a three-way (4*2*2) mixed ANOVA was possible to determine whether there were any significant differences in the ratings of different objects amongst the four groups. However, the assumptions of sphericity and homogeneity of the mixed ANOVA were not met (see Appendix L). Thus, multilevel modelling (MLM) or a linear mixed effects model (LMEM). MLM is a series of regression models that can explain variances at more than one level. It has the benefit of discarding the assumptions of homogeneity and sphericity (Hoffman & Rovine, 2007).

Another reason for adopting MLM is its usefulness when the assumption of interdependence is likely to be violated (Hoffman & Rovine, 2007). MLM is appropriate for nested data (multiple observations are collected from each of the multiple units, Bauer et al., 2020). MLM can explain nested data well because both random and fixed effects can be tested (Field, 2018). Random effects mean assuming the impact of the predictors to be drawn at random from other possibilities. In contrast, fixed effects assume the impact of the predictors are fixed and can represent the whole population of values (Reader, 2019). In this study, two types of random effects (for *objects* and *participants*) were explored via MLM. Firstly, because each participant was presented with 32 objects, each object*individual combination was nested within objects and individuals. A participant's ratings for each object were highly likely to be interrelated. Thus, after considering the effects of ObjectGender and Condition, the effect of different objects on ratings that remained was included as a random effect. Secondly, adding random

effects for participants allowed them to deviate from each other regarding how their responses were affected by ObjectGender or Condition.

In this study, the fixed effects of MLM were *LanguageGroup, ObjectGender* and *Condition*. The random effects included *objects* and *participants*. The MLM was done using R 4.0.4. with the lme4 package (Bates et al., 2015). Data entry was done using Microsoft Excel. To perform MLM, the data were transformed from the wide form (one row per participant with within-subject variables listed horizontally) to the long form (several rows per participant with within-subject variables listed vertically). Six models were compiled linearly, as shown in Table 7. While controlling the random effects for objects and participants, Models 1, 2, and 3 aimed to determine the fixed main effects of LanguageGroup, Condition and ObjectGender. The next three models aimed to identify the fixed interaction effects of LanguageGroup * ObjectGender, LanguageGroup * Condition, and LanguageGroup * Condition * ObjectGender, respectively. To give an example of the codes, Model1 was realised by 'model1 <- lmer(rating ~ LanguageGroup + (1 | Participant) + (1 | Objects), data=mydata[3]' (for all the remaining codes, refer to Appendix J).

Table 7

Multilevel Models

Models:
Model1: rating ~ LanguageGroup + (1 | ID) + (1 | Item)
Model2: rating ~ Condition + (1 | ID) + (1 | Item)
Model3: rating ~ ObjectGender + (1 | ID) + (1 | Item)
Model4: rating ~ LanguageGroup * ObjectGender + (1 | ID) + (1 | Item)
Model5: rating ~ LanguageGroup * Condition + (1 | ID) + (1 | Item)
Model6: rating ~ LanguageGroup * ObjectGender * Condition + (1 | ID) + (1 | Item)

ID refers to participants; Item refers to objects.

The results for the main effects are presented in Table 8. In Model 1, the main effect of LanguageGroup only appeared when comparing the French monolingual group to the English monolingual group (M_{change}=4.356, $p < .05$). In other words, only the two monolingual groups

[3] ~ means the ratings to be predicted by LanguageGroup; (1 | Participant) refers to the random effect; 1 refers to the intercept; the left to the | is the effects that are assumed to be random; the effects should vary across the grouping variable to the right of the | (Blissett, 2017).

differed significantly in the total ratings for objects. No main effect of Condition was found in Model 2. overall, the participants' ratings for the experimental items did not differ significantly from their ratings for the control items (M_{change}=1.629, $p > .05$). Model 3 revealed a main effect of ObjectGender, such that all the participants' ratings for male objects were significantly lower than they were for female objects (M_{change}= -8.621, $p < .05$).

Table 8

Results for Main Effects (Models 1-3)

Parameter	*Model 1*			*Model 2*			*Model 3*		
	Est.	*S.E.*	*t value*	*Est.*	*S.E.*	*t value*	*Est.*	*S.E.*	*t value*
Fixed Effects									
(Intercept-English monolinguals)	47.6	2.1	22.7*						
(Intercept-conditionControl)				48.4	3.4	14.2*			
(Intercept-genderFeminine)							53.8	2.5	21.9*
French monolinguals	4.4	1.5	2.9*						
English-dominant bilinguals	1.3	1.5	0.9						
French-dominant bilinguals	1.8	1.5	1.2						
conditionExperimental				1.6	4.1	0.4			
genderMasculine							-8.6	3.4	-2.5*
Random Effects									
Random object variance (SD)	106.6 (10.3)			109.7 (10.4)			90.5 (9.5)		
Random participants variance (SD)	21.0 (4.5)			22.7 (4.7)			22.7 (4.7)		
Residual variance (SD)	540.3 (23.2)			540.3 (23.2)			540.3 (23.2)		
Fit Statistics									
Maximum Likelihood deviance (number of parameters)	41118 (7)			41127 (5)			41121 (5)		
Akaike Information Criterion	41132			41137			41131		

*p<.05.

Interaction effects are displayed in Table 9. Because two-way and three-way interaction effects can occur within and amongst the four groups, they are only reported briefly in this section; a more detailed explanation of each RQ will be presented later. An interaction effect between LanguageGroup and ObjectGender was found in Model 4. This effect means significant differences in ratings for male/female objects between and within groups. For example, when rating female objects, the French monolingual group differed significantly from the English monolingual group (M_{change}=14.954, $p < .05$), and the French monolingual group's ratings for female objects were significantly higher than they were for male objects (M_{change}=36.149, $p < .05$). Some interaction effects between LanguageGroup and Condition were found in Model 5. For example, the English monolingual group and the French monolingual group differed significantly in the control items (M_{change}=6.311, $p < .05$). However, such a significant difference in control items was not found between the English monolinguals and the English-

dominant bilinguals (M_{change}=2.154, p > .05). Model 6 also found a significant three-way LanguageGroup*Condition*ObjectGender effect. For example, French monolinguals differed significantly from English monolinguals in the rating of female control objects (M_{change}=14.777, p < .05).

Table 9

Results for Interaction Effects (Models 4-6).

Parameter	*Model 4*			*Model 5*			*Model 6*		
	Est.	*S.E.*	*t value*	*Est.*	*S.E.*	*t value*	*Est.*	*S.E.*	*t value*
Fixed Effects									
(Intercept-English monolinguals; genderFeminine)	43.8	2.7	16.3*						
(Intercept-English monolinguals; conditionControl)				45.4	3.6	12.5*			
(Intercept-English monolinguals; genderFeminine; conditionControl							48.8	4.6	10.5*
French monolinguals	14.9	1.8	8.5*	6.3	2.1	3.1*	14.7	2.6	5.5*
English-dominant bilinguals	11.8	1.8	6.7*	2.1	2.1	1.1	6.8	2.6	2.5*
French-dominant bilinguals	12.8	1.8	7.3*	3.2	2.1	1.6	8.7	2.6	3.2*
genderMasculine	7.4	3.6	2.1*				-6.8	6.4	-1.1
French monolinguals: genderMasculine	-21.1	1.9	-11.1*				-16.9	3.4	-4.9*
English-dominant bilinguals: genderMasculine	-21.1	1.9	-10.9*				-9.4	3.4	-2.7*
French-dominant bilinguals: genderMasculine	-22.1	1.9	-11.5*				-10.9	3.4	-3.1*
conditionExperimental				3.1	4.3	0.7	-7.2	5.5	-1.3
French monolinguals: conditionExperimental				-2.8	2.1	-1.3	0.2	2.9	0.1
English-dominant bilinguals: conditionExperimental				-1.2	2.1	-0.6	7.3	2.9	2.5*
French-dominant bilinguals: conditionExperimental				-2.1	2.1	-1.0	5.9	2.9	2.1*
genderMasculine:conditionExperimental							20.7	2.6	2.6*
French monolinguals: genderMasculine: conditionExperimental							-6.2	4.1	-1.4
English-dominant bilinguals: genderMasculine: conditionExperimental							-17.1	4.1	-4.1*
French-dominant bilinguals: genderMasculine: conditionExperimental							-16.1	4.1	-3.8*
Random Effects									
Random object variance (SD)	90.6 (9.5)			109.7 (10.4)			89.5 (9.5)		
Random participants variance (SD)	21.7 (4.6)			21.1 (4.6)			21.8 (4.7)		
Residual variance (SD)	518.1 (22.7)			540.4 (23.2)			515.8 (22.7)		
Fit Statistics									
Maximum Likelihood deviance (number of parameters)	40929 (11)*			41116 (11)			40901 (19)*		
Akaike Information Criterion	40951			41138			40939		

*p<.05.

Model 4 was a significant improvement over Model1 (χ^2 difference (4) = 189, p < .001) and had a smaller Akaike Information Criterion (AIC) value, which indicates a better model (Hoffman & Rovine, 2007). In other words, adding an interaction effect between LanguageGroup and ObjectGender is more suitable for the pure main effects of LanguageGroup. Model 5 was not a significant improvement over Model 1 (χ^2 difference (4) = 2.19, p > .05). This finding indicates that the interaction effect between LanguageGroup and Condition may not be as strong as the interaction between LanguageGroup and ObjectGender.

Model 6 was another significant improvement over Models 4 and 5 (χ^2 difference (8) = 28, $p < .001$; χ^2 difference (8) = 215, $p < .001$). This observation suggests that Model 6, with the three-way interaction effect, was the best model to explain the data. Therefore, the rest of this chapter reports on the results for each RQ based on Model 6.

4.2. Demographic Variables

A multiple regression analysis was conducted to see if the demographic variables (education level, participants' own gender, age, metalinguistic awareness, L2 proficiency, self-perceived L2 proficiency, L2 starting age, L2 exposure, and language qualification) would be significant predictors of the voice ratings of all groups. As these demographic variables were all categorical data, they should be appropriately coded to be entered into a regression model (Field, 2018). Among these demographic variables, participants' gender, L2 starting age, metalinguistic awareness and language qualification were nominal and dummy coded[4]. The other variables were ordinal and were treated as continuous (Williams, 2020). After outliers were winsorised, the assumption of normality was met. Correlation tests were conducted to check the assumption of multicollinearity, which was also satisfied because the demographic variables were not highly correlated ($r<.9$, Field, 2009).

In previous literature, there was often a lack of proficiency testing and the effect of L2 proficiency in linguistic relativity is inconclusive (Bassetti, 2007; Lambelet, 2016). Therefore, the L2 proficiency and L2 exposure were focal variables in the regression. As a result, the regression consisted of two models. As shown in Table 10, in Model 1, education level, participants' own gender, age, metalinguistic awareness, self-perceived L2 proficiency, L2 starting age, and language qualification were entered. Model 2 further included L2 proficiency and L2 exposure to see if Model 2 could explain more variances in the outcome variables.

[4] (male=1, female=0; simultaneous bilingual=1, sequential bilingual=0; high metalinguistic awareness=1, low metalinguistic awareness=0; no language qualification=0, having language qualification=1)

Table 10

Regression Model Summary

Model	*R*	R^2	*Adjusted* R^2	*S.E. of the Estimate*	*Change Statistics*				
					ΔR^2	ΔF	*df1*	*df2*	*Sig.* ΔF
1	.07[a]	0.006	-0.02	199.1	0.006	0.2	7	272	1.0
2	.08[b]	0.007	-0.03	199.8	0.001	0.1	2	270	0.9

a Predictors: (Constant), Language Qualification, Self-perceived Proficiency, Metalinguistic-awareness, Education level, Age, Participants' Own Gender, L2 StartingAge

b Predictors: (Constant), Language Qualification, Self-perceived Proficiency, Metalinguistic-awareness, Education level, Age, Participants' Own Gender, L2 StartingAge, L2 Exposure, L2 Proficiency

c Dependent Variable: Response

Table 11

Regression: ANOVA Table

Model		*Sum of Squares*	*df*	*Mean Square*	*F*	*Sig.*
1	Regression	68290.6	7	9755.8	0.2	.9[b]
	Residual	10781809.3	272	39639.1		
	Total	10850099.9	279			
2	Regression	73982.2	9	8220.2	0.2	.9[c]
	Residual	10776117.7	270	39911.5		
	Total	10850099.9	279			

From Table 10, it was apparent that Model 1 did not explain a significant proportion of variances (ΔR^2=0.6%, p>.05). Model 2 only explained slightly more variances than Model 1 (ΔR^2=0.1%, p>.05). Table 11 showed that neither of the models could explain the variances significantly (p>.05). In other words, none of the demographic variables was significant predictors to participants' responses.

Table 12 further showed how many variances could be predicted by these demographic variables.

Table 12

Regression Coefficients

Model		*Unstandardized Coefficients*		*Standardized Coefficients*	*t*	*Sig.*	*95.0% CI for B*		*Correlations*			*Collinearity Statistics*	
		B	*S.E.*	*Beta*			*Lower*	*Upper*	*Zero-order*	*Partial*	*Part*	*Tolerance*	*VIF*
1	(Constant)	317.7	113.4		2.8	0.0	94.5	540.9					
	OwnGender	0.6	27.9	0.0	0.0	1.0	-54.3	55.6	0.0	0.0	0.0	0.7	1.4
	Age	1.5	17.9	0.0	0.1	0.9	-33.7	36.6	0.0	0.0	0.0	0.9	1.2
	EduLevel	17.9	20.7	0.1	0.9	0.4	-22.9	58.6	0.1	0.1	0.1	0.9	1.1
	SelfPerceProf	6.5	16.4	0.0	0.4	0.7	-25.8	38.8	0.0	0.0	0.0	0.8	1.2
	MetaLinAwa	-10.8	26.2	0.0	-0.4	0.7	-62.3	40.7	0.0	0.0	0.0	0.9	1.2
	StartingAge	13.1	40.3	0.0	0.3	0.7	-66.2	92.5	0.0	0.0	0.0	0.7	1.4
	Qualifications	0.3	28.1	0.0	0.0	1.0	-55.0	55.6	0.0	0.0	0.0	0.7	1.4
2	(Constant)	320.0	114.3		2.8	0.0	95.0	545.1					
	OwnGender	1.6	28.3	0.0	0.1	1.0	-54.1	57.2	0.0	0.0	0.0	0.7	1.4
	Age	1.7	18.0	0.0	0.1	0.9	-33.8	37.2	0.0	0.0	0.0	0.9	1.2
	EduLevel	16.8	21.3	0.1	0.8	0.4	-25.2	58.7	0.1	0.0	0.0	0.8	1.2
	SelfPerceProf	7.0	16.9	0.0	0.4	0.7	-26.2	40.2	0.0	0.0	0.0	0.8	1.3
	MetaLinAwa	-15.8	29.8	0.0	-0.5	0.6	-74.5	42.9	0.0	0.0	0.0	0.7	1.5
	StartingAge	7.6	44.0	0.0	0.2	0.9	-79.0	94.3	0.0	0.0	0.0	0.6	1.7
	Qualifications	4.0	30.9	0.0	0.1	0.9	-56.7	64.8	0.0	0.0	0.0	0.6	1.7
	L2Proficiency	2.7	7.5	0.0	0.4	0.7	-12.0	17.4	0.0	0.0	0.0	0.6	1.7
	L2Exposure	-0.2	10.0	0.0	0.0	1.0	-19.7	19.4	0.0	0.0	0.0	0.6	1.6

The column unstandardised *B* shows the change in the voice ratings associated with a unit change in each demographic predictor (Field, 2018). The *t*-statistic and its associated *p*-value in each row represented if the change (*b*-value) was significantly different. As the *p*-value of each predictor was more prominent than 0.05, the null hypotheses could not be rejected. Thus, no demographic variables were significant predictors of the overall ratings.

The proficiency tests in this study were crucial because significant differences were found between participants' self-perceived L2 proficiency and their actual L2 proficiency. Two Wilcoxon signed-rank tests indicate that English-dominant bilinguals' self-perceived L2 proficiencies were significantly higher than their tested French proficiency ($Z = -4.756$, $p < .001$). By contrast, French-dominant bilinguals' self-perceived English proficiencies were significantly lower than their actual capabilities ($Z = -2.0$, $p < .05$).

Another intriguing issue is whether simultaneous and sequential bilinguals differ in their ratings. As most sequential bilinguals recruited in this study did not start learning L2 until ten, they could potentially show different ratings compared to simultaneous ones. Another several independent samples *t*-tests were conducted for the English-dominant and French-dominant bilingual groups. Results (Table 13 and Table 14)

showed that sequential and simultaneous bilinguals did not differ significantly in their ratings for each Condition*ObjectGender combination (experimental masculine, experimental feminine, control masculine, and control feminine objects).

Table 13

Simultaneous Bilinguals vs Sequential Bilinguals: Descriptives

	English-dominant Bilinguals	N	M	S.D.	S.E. Mean
Experimental Masculine	Sequential	26	514.2	107.6	21.1
	Simultaneous	9	366.6	191.8	63.9
Experimental Feminine	Sequential	26	592.3	158.5	31.1
	Simultaneous	9	676.2	259.5	86.5
Control Masculine	Sequential	26	209.0	61.8	12.1
	Simultaneous	9	163.6	65.3	21.8
Control Feminine	Sequential	26	262.0	51.4	10.1
	Simultaneous	9	326.4	98.9	33.0
	French-dominant Bilinguals	N	M	S.D.	S.E. Mean
Experimental Masculine	Sequential	34	472.6	115.8	19.9
	Simultaneous	1	558.0	.	.
Experimental Feminine	Sequential	34	625.3	143.2	24.6
	Simultaneous	1	419.0	.	.
Control Masculine	Sequential	34	200.4	62.8	10.8
	Simultaneous	1	159.0	.	.
Control Feminine	Sequential	34	287.8	64.9	11.1
	Simultaneous	1	293.0	.	.

Table 14

Simultaneous Bilinguals vs Sequential Bilinguals: Independent Samples T-tests

English-dominant Bilinguals	*t*	*df*	*Sig.*	*M Difference*	*S.D. Difference*	*95% CI of the Difference*	
			Two-Sided p			*Lower*	*Upper*
Experimental Masculine	2.2	9.8	0.1	147.7	67.3	-2.8	298.1
Experimental Feminine	-0.9	10.1	0.4	-83.9	91.9	-288.3	120.5
Control Masculine	1.9	33.0	0.1	45.5	24.2	-3.8	94.8

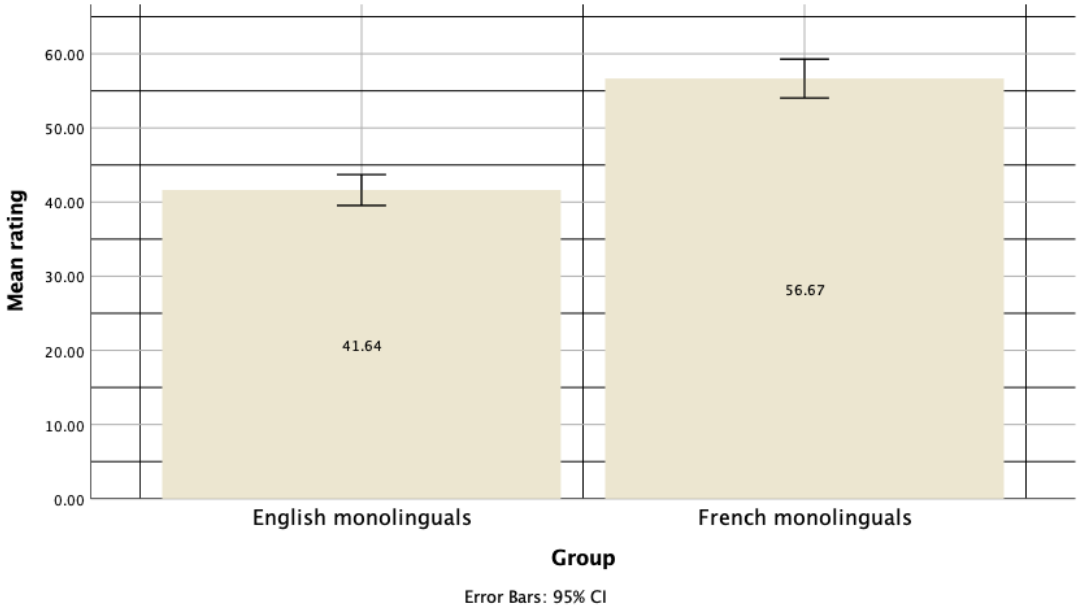
Control Feminine	-1.9	9.5	0.1	-64.4	34.5	-141.8	12.9
French-dominant Bilinguals	*t*	*df*	*Sig. Two-Sided p*	*M Difference*	*S.D. Difference*	*95% CI of the Difference*	
						Lower	*Upper*
Experimental Masculine	-0.7	33.0	0.5	-85.4	117.5	-324.5	153.8
Experimental Feminine	1.4	33.0	0.2	206.3	145.3	-89.4	502.0
Control Masculine	0.7	33.0	0.5	41.4	63.8	-88.3	171.1
Control Feminine	-0.1	33.0	0.9	-5.2	65.9	-139.3	128.8

4.3. English Monolinguals versus French Monolinguals

The ratings of 35 English and 35 French monolinguals for objects of four Condition*ObjectGender trials are presented in Graph 2, 3, 4, and 5.

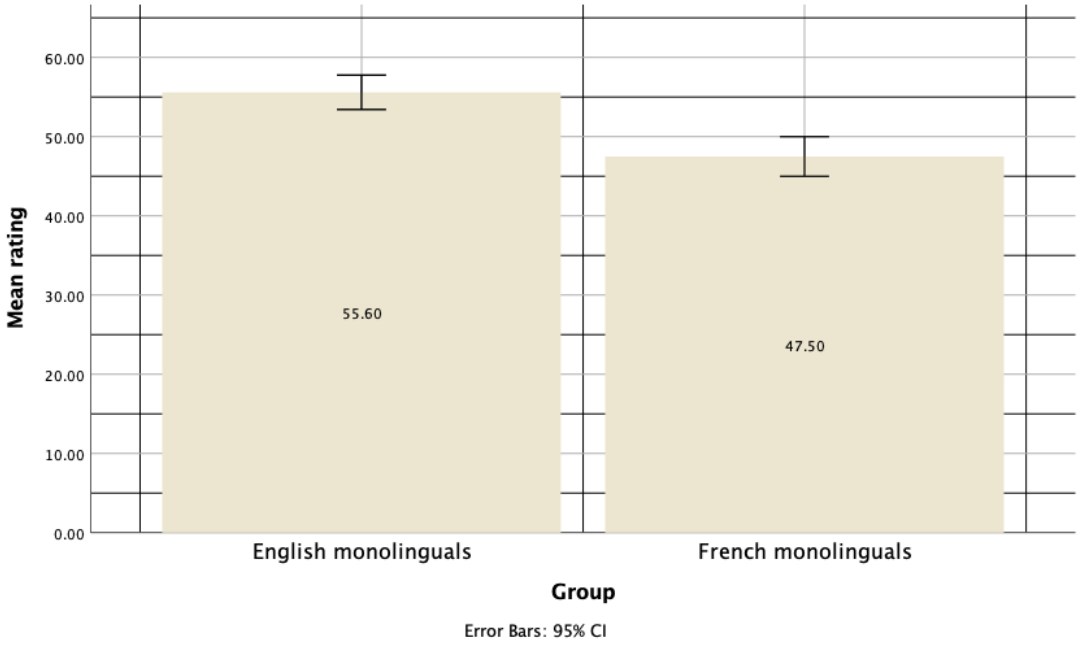
Graph 2

English versus French Monolinguals: Mean Rating of Experimental Feminine Objects



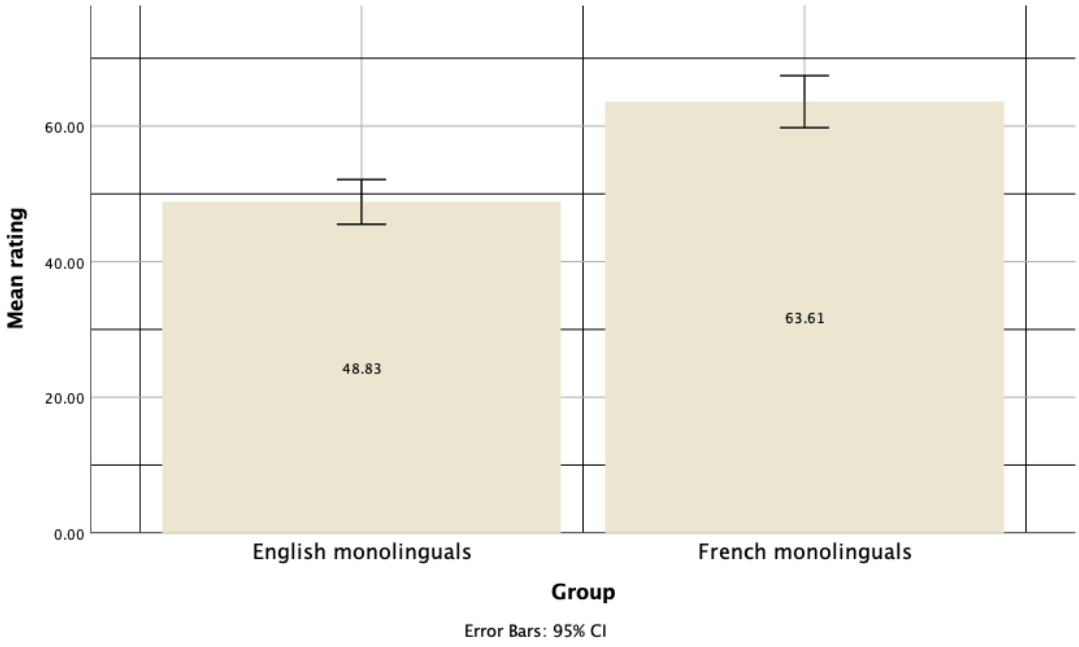
Graph 3

English versus French Monolinguals: Mean Rating of Experimental Masculine Objects



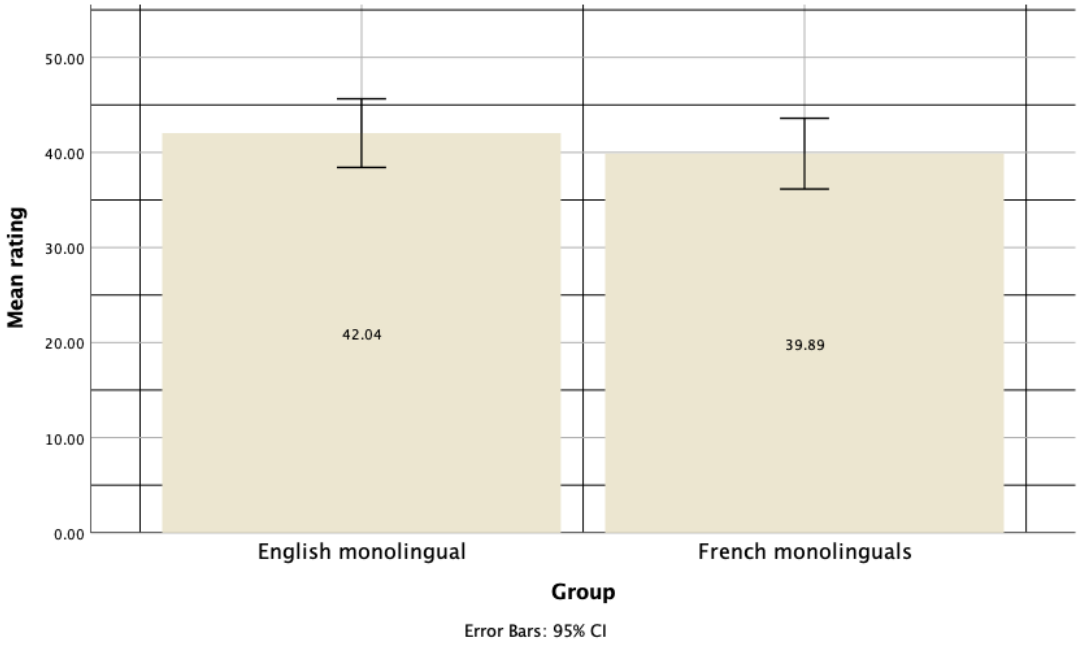
Graph 4

English versus French Monolinguals: Mean Rating of Control Feminine Objects



Graph 5

English versus French Monolinguals: Mean Rating of Control Masculine Objects



The comparison between English and French monolinguals was conducted based on the most appropriate model, Model 6 with the three-way interaction. Four independent sample *t*-tests were conducted to compare the ratings in each Condition (2) * ObjectGender (2) trial, as shown in Table 15.

Table 15

***T*-tests Results for English and French Monolinguals**

Trial	*Group*	*M*	*S.D.*	*S.E.M*	*df*	*t*	*p (2-tailed)*	*95% CI*	*d*
Experimental*Female	Eng. Mono	41.64	20.80	1.06	731	-8.8	<.01	-18.4, -11.7	-.6
	Fren. Mono	56.67	26.16	1.33					
Experimental*Male	Eng. Mono	55.60	21.65	1.10	753	4.8	<.01	4.8, 11.4	.3
	Fren. Mono	47.50	24.94	1.27					
Control*Female	Eng. Mono	48.83	22.14	1.67	340	-5.8	<.01	-19.8, -9.7	-.6
	Fren. Mono	63.61	25.73	1.95					
Control*Male	Eng. Mono	42.04	24.19	1.89	348	.8	.41	-3.0, 7.3	.1
	Fren. Mono	39.89	24.96	1.89					

The results of the independent sample *t*-tests indicated that there was a significant difference in the English and French monolingual participants' ratings for the experimental female items: $t(731) = -8.8$, $p<.01$. In other words, the French monolinguals rated the experimental female items significantly higher ($M = 56.67$, $SD = 26.16$) than did the English monolinguals ($M = 41.64$, $SD = 20.80$). Similarly, there was a significant difference for experimental male items

between the two groups: t (753) = 4.8, p <.01. The French monolinguals also rated experimental male items significantly lower (M = 47.50, SD = 24.94) than did the English monolingual participants (M = 55.60, SD = 21.65). The higher rating for experimental female items and a lower rating for experimental male items indicated the effect of the French grammatical gender system on French monolinguals; in other words, they consistently rated the objects following the objects' gender in French.

A note of caution is due here since multiple comparisons of the same sample of data increase the bias of making a Type I error or falsely rejecting the H_0 (Mittelhammer et al., 2000). Therefore, a Bonferroni correction was applied to reduce the bias by dividing the alpha threshold (here, p=.05) by the number of comparisons made (here, 4) to have a revised alpha threshold that is stricter (revised p =.0125). The independent samples t-tests were still significant after the Bonferroni correction was applied (p<.0125).

With regards to the control items, there was a significant difference for control female items: t (340) = -5.8, p<.01. French monolinguals rated control female items significantly higher (M = 63.61, SD = 25.73) than did English monolinguals (M = 48.83, SD = 22.14). However, no significant difference was found for control male items: t (348) = .8, p > .05.

A follow-up chi-square test was conducted to determine whether the groups' open-ended answers regarding the ten items rated highly differently by English and French monolinguals in the pilot study differed. The qualitative responses were coded according to three levels (1=grammatical gender system, 2=object's characteristics or feelings, 3=other). A significant association was found between the groups and the answers: χ^2 (2) = 92.7, p < .01. French monolinguals gave answers that mapped to the language's grammatical gender system significantly more often (N = 82) than did English monolinguals (N = 2). English monolinguals relied more on the objects' characteristics to rate their genders (N = 307) than French monolinguals (N = 217). This finding provided qualitative evidence regarding the effect of the

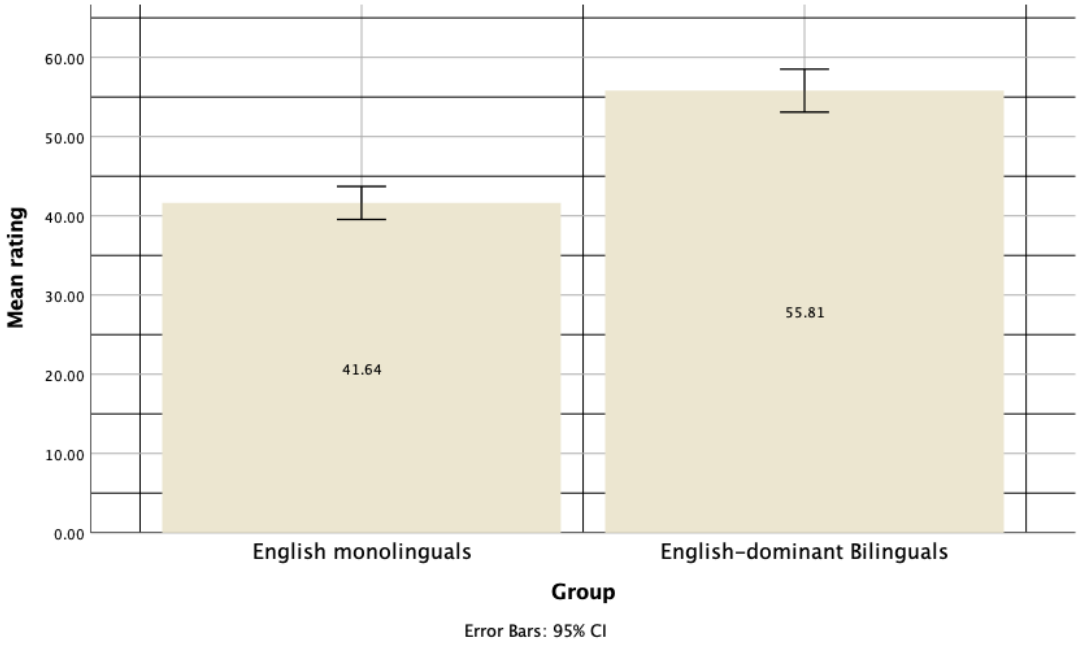
French gender system. French monolinguals explicitly claimed that they rated each object according to its gender in the French language.

4.4. English Monolinguals versus English-dominant Bilinguals

The ratings of 35 English and 35 English-dominant bilinguals for objects of four Condition*ObjectGender trials are presented in Graph 6, 7, 8, and 9.

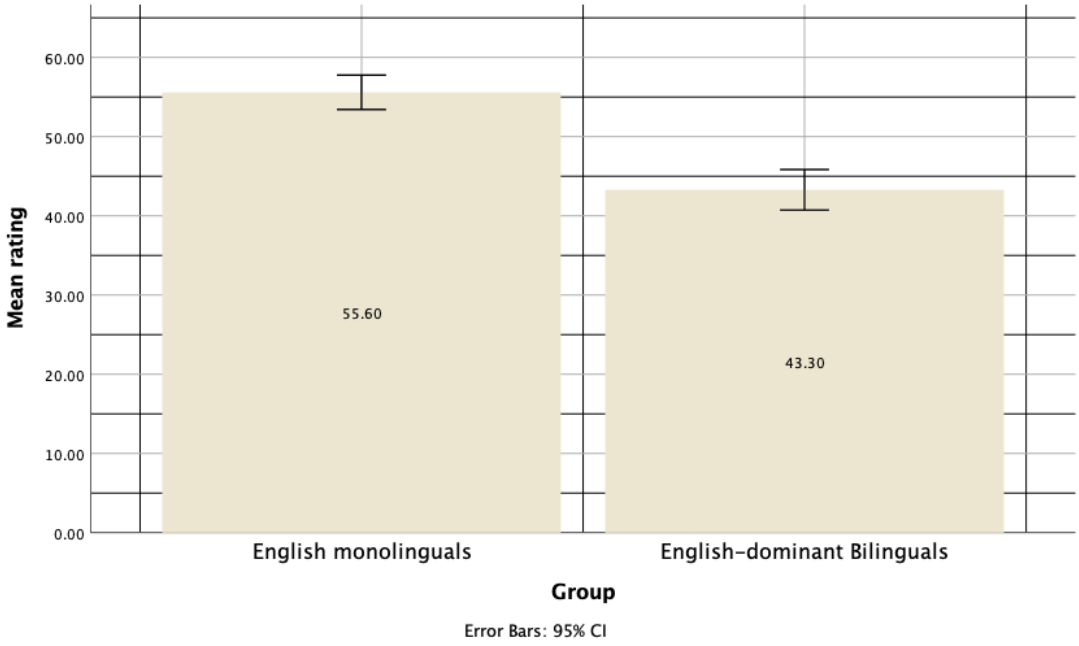
Graph 6

English Monolinguals versus English-dominant Bilinguals: Mean Rating of Experimental Feminine Objects



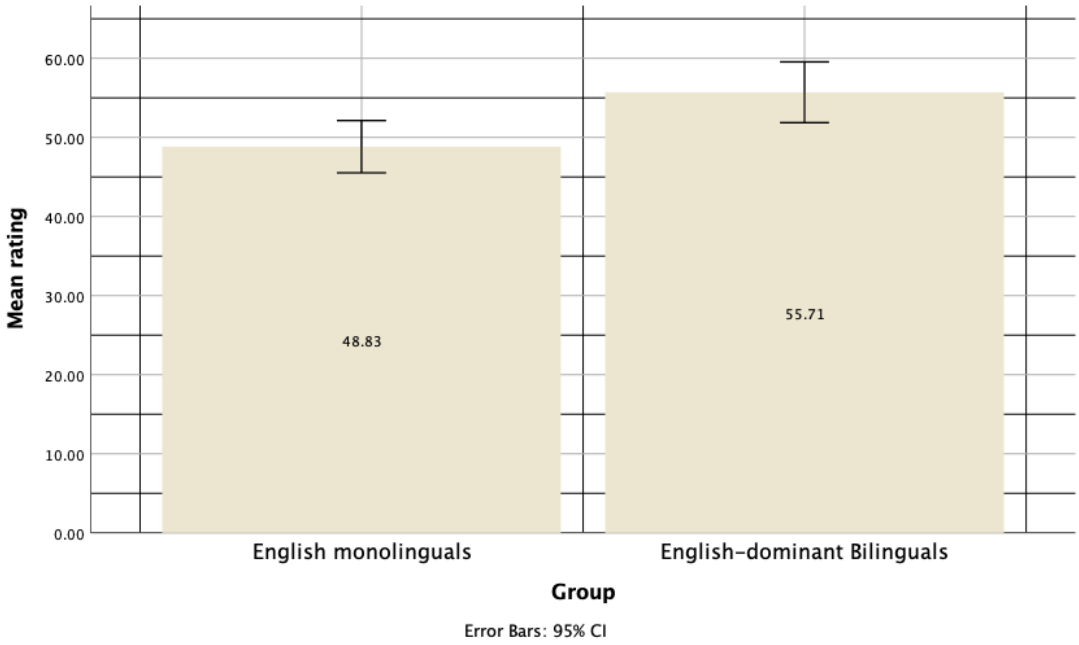
Graph 7

English Monolinguals versus English-dominant Bilinguals: Mean Rating of Experimental Masculine Objects



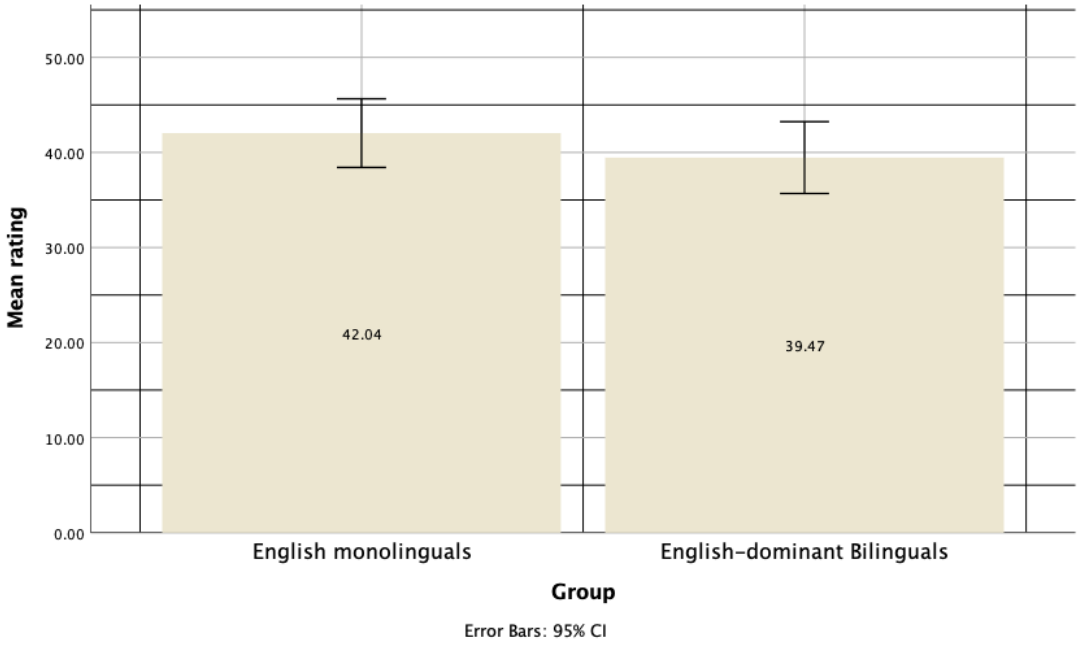
Graph 8

English Monolinguals versus English-dominant Bilinguals: Mean Rating of Control Feminine Objects



Graph 9

English Monolinguals versus English-dominant Bilinguals: Mean Rating of Control Control Objects



Based on Model 6, four independent sample *t*-tests were conducted (with a Bonferroni correction applied) to compare the ratings in each Condition (2) * ObjectGender (2) trial, as shown in Table 16.

Table 16

***T*-tests Results for English Monolinguals and English-dominant Bilinguals**

Trial	*Group*	*M*	*S.D.*	*S.E.M*	*df*	*t*	*p (2-tailed)*	*95% CI*	*d*
Experimental*Female	Eng. Mono	41.64	20.80	1.06	720	-8.1	<.01	-17.6, -10.8	-.6
	Eng-dom. Bil	55.81	27.07	1.38					
Experimental*Male	Eng. Mono	55.60	21.65	1.10	748	7.2	<.01	8.9, 15.7	.5
	Eng-dom. Bil	43.29	25.56	1.30					
Control*Female	Eng. Mono	48.83	22.14	1.67	340	-2.7	<.01	-11.9, -1.8	-.3
	Eng-dom. Bil	55.71	25.77	1.95					
Control*Male	Eng. Mono	42.04	24.19	1.83	348	1.0	.33	-2.6, 7.8	.1
	Eng-dom. Bil	39.47	25.32	1.91					

The results of the independent sample *t*-tests indicated that there was a significant difference in the ratings of the English monolinguals and the English-dominant bilinguals for experimental female items: t (720) = -8.1, p<.01. In other words, the English-dominant bilinguals rated experimental female items significantly higher (M = 55.81, SD = 27.07) than did the English monolinguals (M = 41.64, SD = 20.80). Similarly, there was a significant difference for the experimental male items between the two groups: t (748) = 7.2, p <.01. The English-dominant

bilinguals also rated experimental male items significantly lower (M = 43.29, SD = 25.56) than did the English monolinguals (M = 55.60, SD = 21.65). A higher rating for experimental female items and a lower rating for experimental male items indicated the effect of the French grammatical gender system on English-dominant bilinguals. In other words, the English-dominant bilinguals' perceptions were affected by their L2 because of having acquired the French grammatical gender system.

Regarding the control items, there was a significant difference for the control female items: t (340) = -2.7, p<.01. The English-dominant bilinguals rated control female items significantly higher (M = 55.71, SD = 25.77) than did the English monolinguals (M = 48.83, SD = 22.14). No significant difference was found for control male items: t (348) = 1.0, p > .05. The findings were in consonance with those for the French monolinguals in this case. This might indicate that the French grammatical gender system can increase the tendency to assign a feminine gender to objects.

Follow-up correlation tests revealed no significant association between English-dominant bilinguals' French proficiency and their ratings for experimental female (r = .36, p>.05), experimental male (r=.16, p>.05) or control female items (r=.34, p>.05). This means that a higher level of French proficiency does not indicate a more potent effect of the French grammatical gender system. Similarly, no significant association was found between ratings and L2 exposure, nor between ratings and metalinguistic awareness.

A follow-up chi-square test revealed a significant association between the groups and their answers: χ^2 (2) = 88.9, p < .001. Like the French monolinguals, the English-dominant bilinguals answered significantly more in accord with the language' s grammatical gender system (N = 83) than the English monolinguals (N = 2). The English monolinguals relied more (N = 307) on the objects' characteristics to rate their genders than the French monolinguals (N = 228). This finding also supports the effect of language on the perception of speakers of a genderless

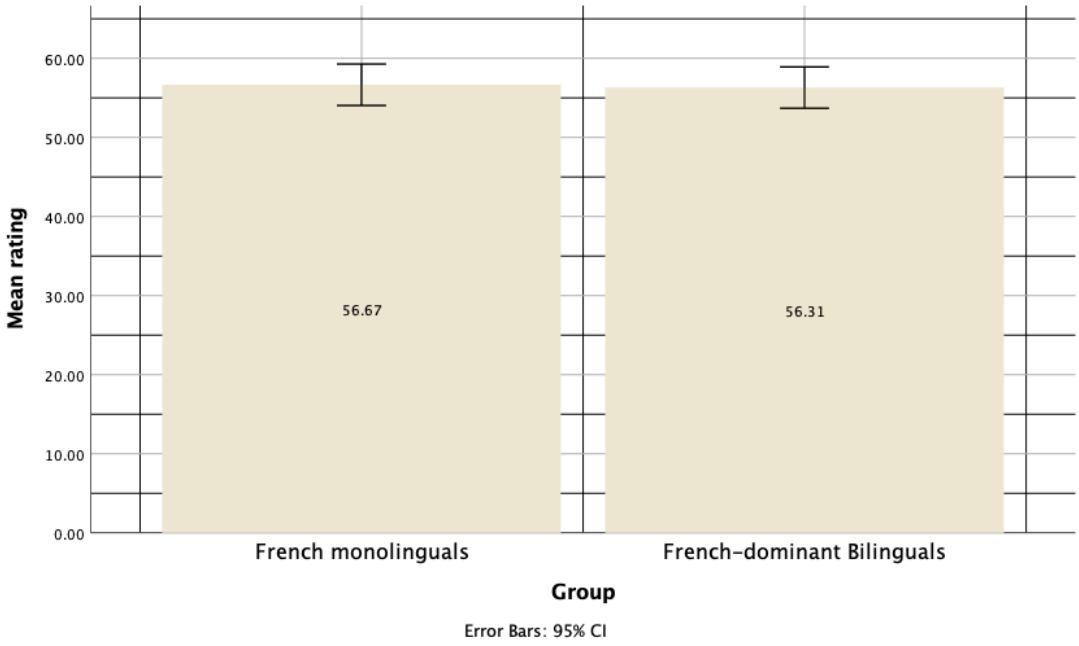
language due to the acquisition of the French gender system.

4.5. French Monolinguals versus French-dominant Bilinguals

The ratings of 35 French monolinguals and 35 French-dominant bilinguals for objects of four Condition*ObjectGender trials are presented in Graph 10, 11, 12, and 13.

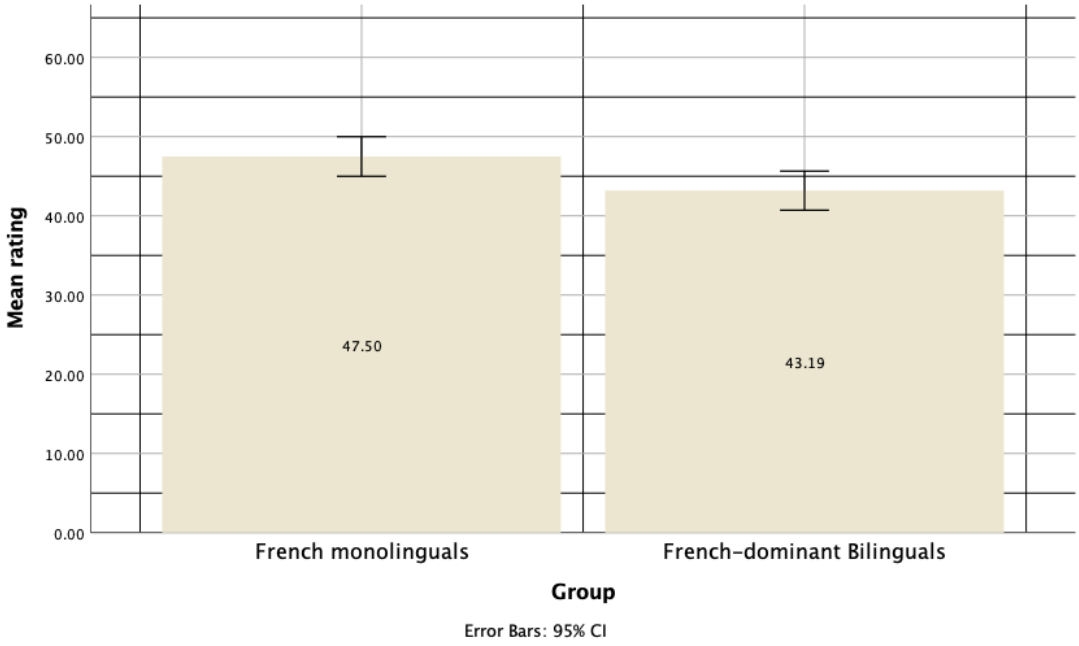
Graph 10

French Monolinguals versus French-dominant Bilinguals: Mean Rating of Experimental Feminine Objects



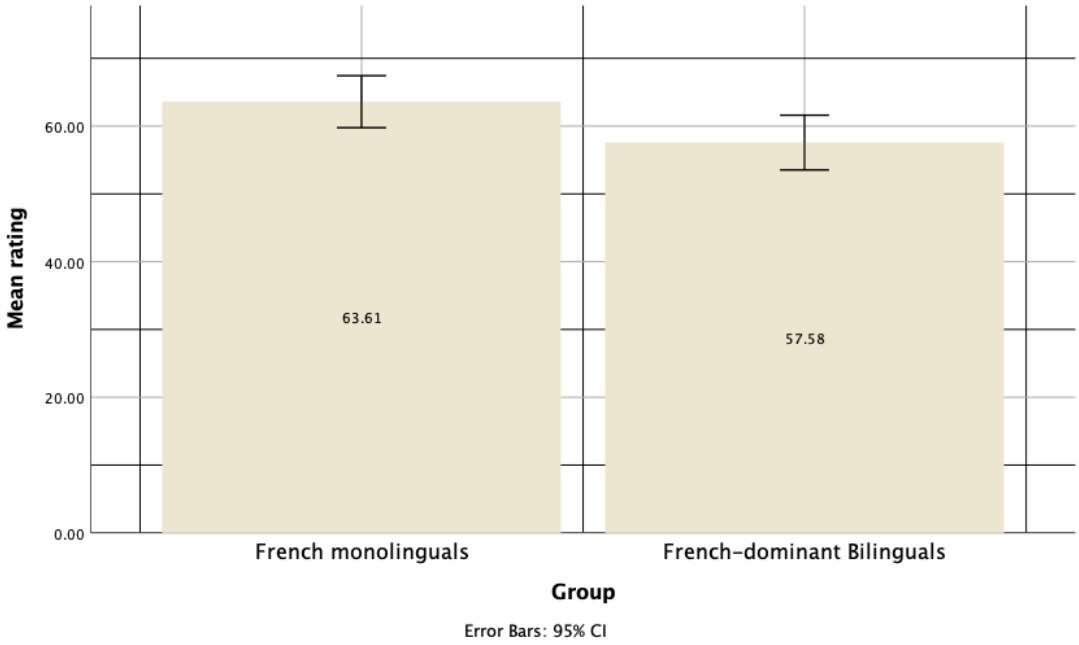
Graph 11

French Monolinguals versus French-dominant Bilinguals: Mean Rating of Experimental Masculine Objects



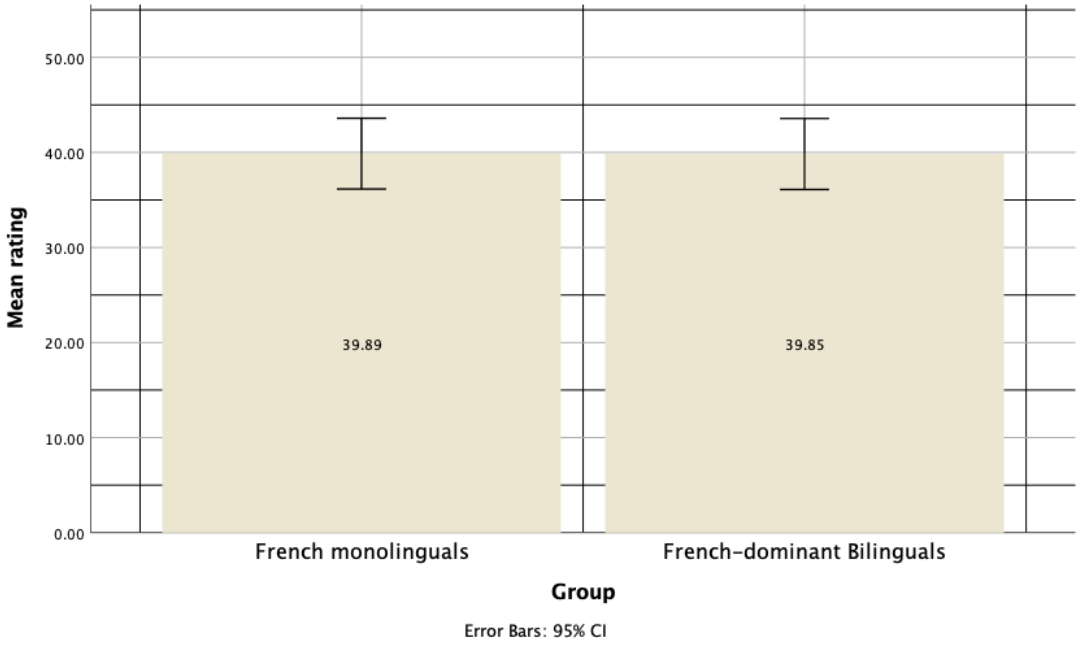
Graph 12

French Monolinguals versus French-dominant Bilinguals: Mean Rating of Control Feminine Objects



Graph 13

French Monolinguals versus French-dominant Bilinguals: Mean Rating of Control Masculine Objects



Based on Model 6, four independent sample *t*-tests were conducted (with a Bonferroni correction applied) to compare the ratings in each Condition (2) * ObjectGender (2) trial, as shown in Table 17.

Table 17

***T*-tests Results for French Monolinguals and French-dominant Bilinguals**

Trial	*Group*	*M*	*S.D.*	*S.E.M*	*df*	*t*	*p (2-tailed)*	*95% CI*	*d*
Experimental*Female	Fre. Mono	56.67	26.16	1.33	768	.2	.85	-3.3, 4.1	.01
	Fre-dom. Bil	56.31	26.10	1.33					
Experimental*Male	Fre. Mono	47.50	24.94	1.27	768	2.4	.02	.8, 7.8	.2
	Fre-dom. Bil	43.18	24.62	1.25					
Control*Female	Fre. Mono	63.61	25.73	1.95	348	2.1	.03	.5, 11.6	.2
	Fre-dom. Bil	57.58	27.10	2.05					
Control*Male	Fre. Mono	39.89	24.96	1.89	348	1.0	.99	-5.2, 5.3	.0
	Fre-dom. Bil	39.85	25.02	1.89					

The results of the independent sample *t*-tests indicated that there was no significant difference in the ratings for experimental female objects: t (768) = .2, p>.05. In other words, French-dominant bilinguals rated experimental female items similarly (M = 56.31, SD = 26.1) to French monolinguals (M = 56.67, SD = 26.16). There was also no significant difference for experimental male items between the two groups after the Bonferroni correction was employed: t (768) = 2.4, p =.02 > .0125. French-dominant bilinguals only rated experimental male items

insignificantly lower (M = 43.18, SD = 24.62) than did French monolinguals (M = 47.50, SD = 24.94). With regards to the control items, there was no significant difference for control female items: t (348) = 2.1, p=.03 > .0125. French-dominant bilinguals rated control female items insignificantly lower (M = 57.58, SD = 27.09) than did French monolinguals (M = 63.61, SD = 25.73). No significant difference was found for control male items: t (348) = 1.0, p > .05. These findings might indicate that learning a genderless language does not significantly weaken gendered perceptions of specific objects.

No significant association was found between French-dominant bilinguals' French proficiency and their ratings (r=-.18, p>.05). This means that a higher level of English proficiency may not contribute to weakening the effect of the French grammatical gender system. Other demographic variables (metalinguistic awareness, L2 exposure, starting age of L2) were not significantly correlated with the ratings.

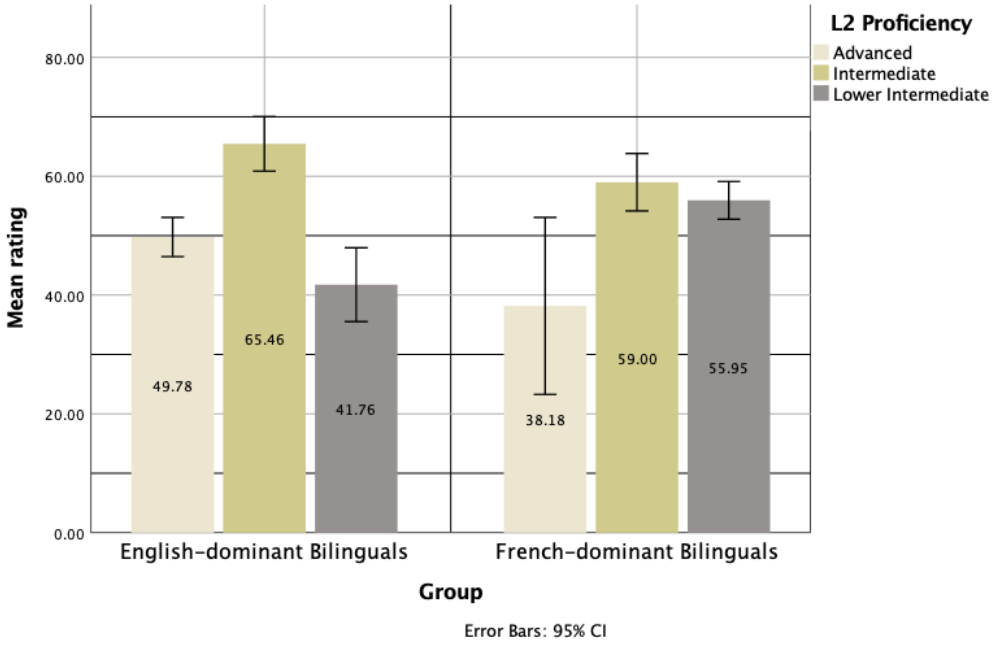
A follow-up chi-square test revealed that there was no significant relationship between the groups and the answers: χ^2 (2) = 1.4, p > .05. French monolinguals and French-dominant bilinguals produced similar answer patterns regarding the grammatical gender system (N=82; N=76), objects' characteristics (N=217; N=231) and other features (N=51; N=43). This finding again might provide qualitative evidence that the acquisition of a genderless language might not diminish the effect of the French grammatical gender system.

4.6. English-dominant Bilinguals vs French-Dominant Bilinguals

The ratings of 35 English- and 35 French-dominant Bilinguals for objects of four Condition*ObjectGender trials are presented in Graph 14, 15, 16, and 17.

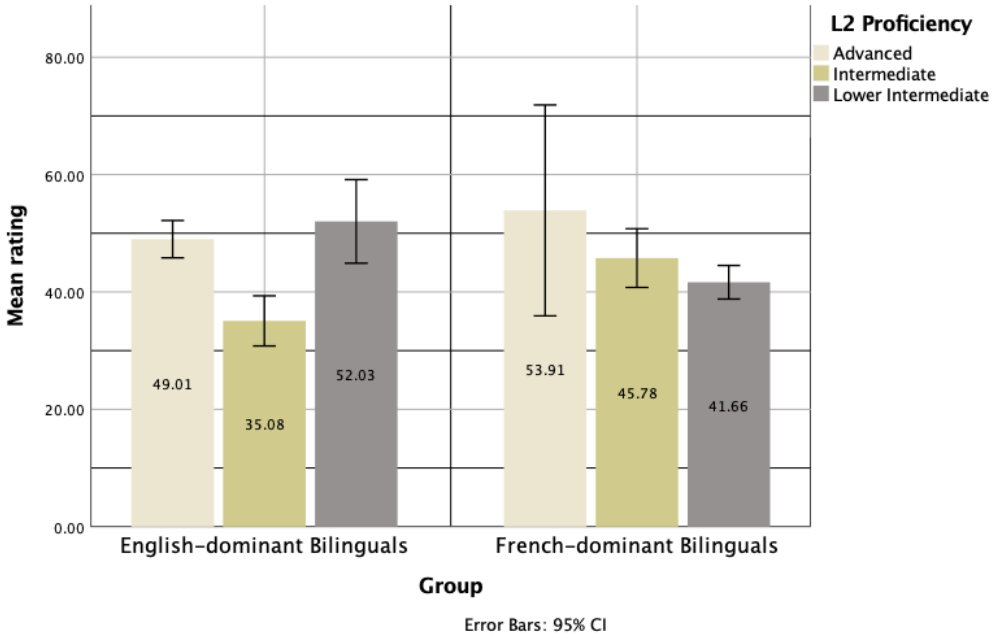
Graph 14

French-dominant Bilinguals v.s. French-dominant Bilinguals: Mean Rating of Experimental Feminine Objects by L2 Proficiency



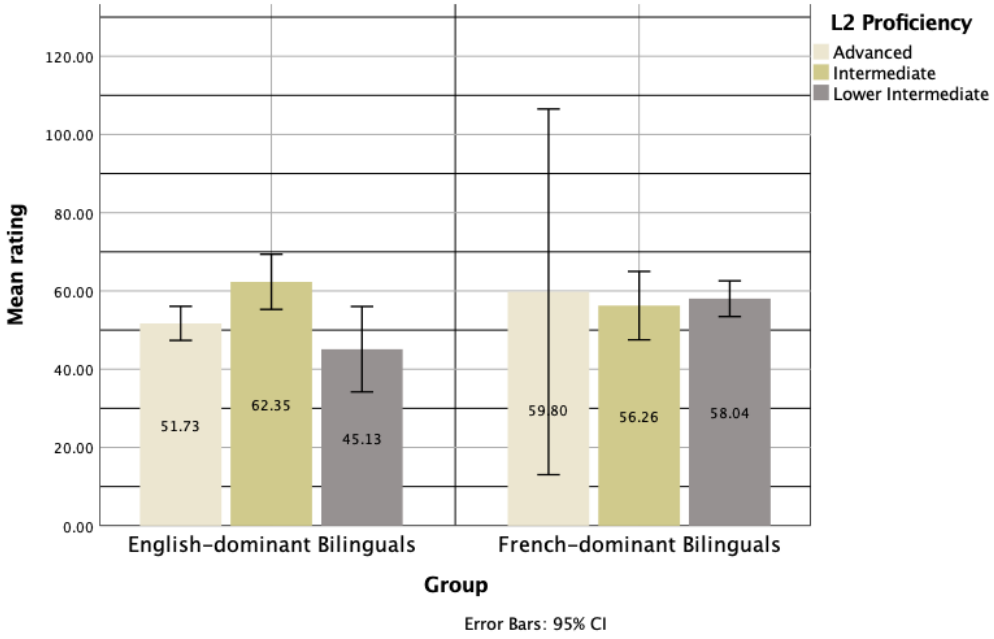
Graph 15

French-dominant Bilinguals v.s. French-dominant Bilinguals: Mean Rating of Experimental Masculine Objects by L2 Proficiency



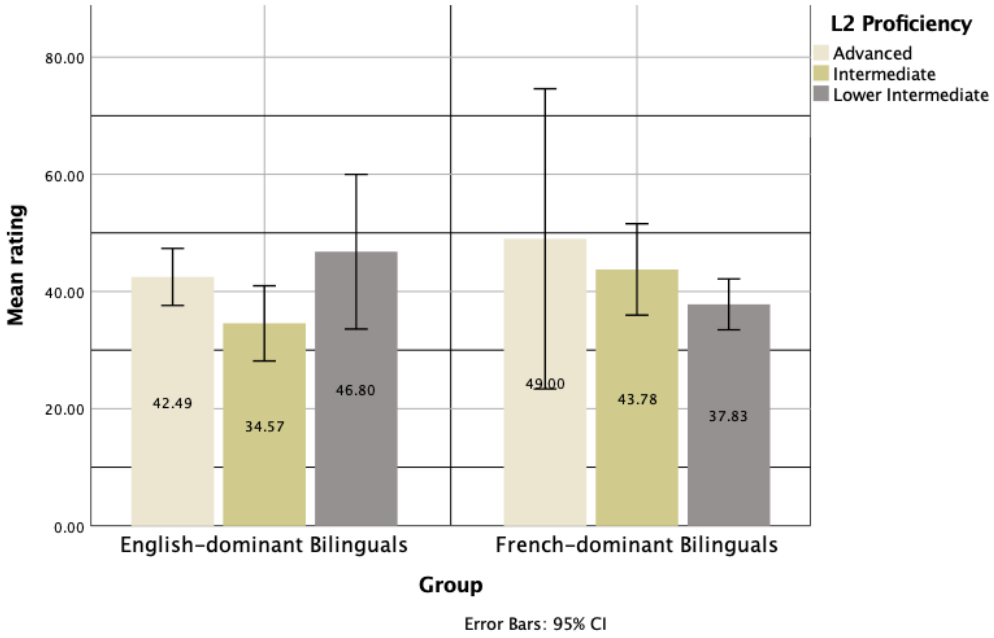
Graph 16

French-dominant Bilinguals v.s. French-dominant Bilinguals: Mean Rating of Control Feminine Objects by L2 Proficiency



Graph 17

French-dominant Bilinguals v.s. French-dominant Bilinguals: Mean Rating of Control Masculine Objects by L2 Proficiency



Based on Model 6, four independent sample t-tests were done to compare ratings in each Condition (2) * ObjectGender (2) trial, as shown in Table 18.

Table 18

***T*-tests Results for English-dominant Bilinguals and French-dominant Bilinguals**

Trial	*Group*	*M*	*S.D.*	*S.E.M*	*df*	*t*	*p (2-tailed)*	*95% CI*	*d*
Experimental*Female	Eng-dom. Bil	55.81	27.07	1.38	768	-.26	.79	-4.3, 3.3	-.02

	Fre-dom. Bil	56.31	26.10	1.33					
Experimental*Male	Eng-dom. Bil	43.29	25.53	1.30	768	.06	.95	-3.4, 3.7	.00
	Fre-dom. Bil	43.18	24.62	1.25					
Control*Female	Eng-dom. Bil	55.71	25.77	1.95	348	-.66	.51	-7.4, 3.7	-.07
	Fre-dom. Bil	57.58	27.10	2.05					
Control*Male	Eng-dom. Bil	39.47	25.31	1.91	348	-.14	.89	-5.7, 4.9	-.02
	Fre-dom. Bil	39.85	25.02	1.89					

Results of independent samples t-tests revealed no significant differences in the ratings for any Condition*ObjectGender combination. In other words, the two bilingual groups rated items similarly. However, there was a strong imbalance in the participants' L2 proficiency, χ^2 (2) = 31.5, $p < .001$. French participants with high English proficiency (N=24) significantly outnumbered English advanced speakers of French (N=3). On the contrary, English participants with low French proficiency (N=17) were significantly more than French participants with a low level of English (N=1). Thus, an ANCOVA with multiple covariates (L2 proficiency, metalinguistic awareness, L2 exposure, starting age of L2) in each Condition*ObjectGender trial was done, as the assumptions of normality and homogeneity were met. The results showed that, after controlling for covariates, there was a significant difference only in the ratings for control female objects ($F(1, 63)=10.57$, $p<.05$) between the two groups. French-dominant bilinguals rated control female items slightly higher than English-dominant bilinguals when the effect of proficiency-related covariates was controlled.

A follow-up chi-square test showed no significant relationship between group and answers, χ^2 (2) = .52, $p > .05$. English-dominant bilinguals and French-dominant bilinguals produced similar answer patterns in the grammatical gender system (N=83; N=76), in objects' characteristics (N=228; N=231), and other answers (N=39; N=43).

5. Discussion

This study investigated linguistic relativity, specifically, the effect of the grammatical gender system of French (both as L1 and L2) on its speakers' and learners' perceptions of objects. Firstly, the study investigated if (and if so, how) the grammatical gender system in French affects French monolinguals compared with English monolingual speakers. Secondly, it explored if (and if so, how) learning a gendered or genderless L2 affects bilingual adults' perception of objects compared to monolingual speakers. Thirdly, it tested if the patterns of object perception between English-dominant bilingual adults and French-dominant bilingual adults are different and, if so, how.

The findings of the study showed that there was an effect of the French grammatical gender system. The results indicated that the French monolinguals and the English-dominant bilinguals assigned objects' gender consistently with the French grammatical gender system, in contrast to English monolinguals. Secondly, the effect of French grammatical gender was not significantly affected by the acquisition of a genderless L2 (English). This was showcased in the result that French-dominant bilinguals did not differ significantly from French monolinguals in ratings, though having lower ratings for experimental masculine and control feminine objects in descriptive statistics. This chapter discusses the findings with theories, previous results, pedagogical implications, and areas for improvement.

5.1. The Effect of the French Grammatical Gender System on Perception

The results of this experiment support linguistic relativity in that an effect of the French grammatical gender system did exist. Concerning the first two research questions, it was found that both French monolinguals and English-dominant bilinguals assigned gender to objects consistently with the objects' French gender. The findings support the predictions (the effect of the French gender system as both L1 and L2). Additionally, in their open-ended response, both groups used significantly more reasons concerning grammatical gender to explain their choices. By contrast, English monolinguals randomly assigned gender to experimental objects

based on the object's characteristics and properties. The difference between English monolinguals and French monolinguals and English monolinguals and English-dominant bilinguals shows the effect of French on its speakers' perception, both as an L1 and L2. Also, because there was no significant difference in sequential and simultaneous bilinguals in ratings, the interpretation can thus be argued to hold for both early and sequential bilinguals. It is important to note that the effect of French as an L2 on perception can only be interpreted among speakers with a genderless L1.

The results reflect the findings of previous studies in the subject area in question (e.g., Bassetti, 2007). Further, the study attempts to improve previous research design in the subject area by testing L2 proficiency (participants needed to reach an L2 vocabulary score of 60% to be deemed bilingual) and setting control items based on the pilot to diminish the confounding effect of gender stereotypes in objects. Thus, the study's conclusions can be relatively reliable that the impact of the grammatical gender system holds.

Such an effect of the French grammatical gender system can have practical societal implications. For example, children's literature has been found to present anthropomorphised objects or animals with genders following the grammatical gender system (Bassetti & Nicoladis, 2016; Mills, 1986). These gendered figures could influence children's perceptions and impose gender biases from a very young age. This idea was supported by the responses of the French monolinguals and French-dominant bilinguals in the experiment's explanation questions. When assigning gender to *elephant*, most French participants considered it masculine, following the elephant's gender in French. However, four participants explained that their score was motivated by recalling Babar, a fictional elephant character in the French children's book *Babar the Elephant*. This explanation shows the long-lasting effect of character depictions in children's literature on their perception. Such an effect also appears in children's production. A recent study conducted by Hsiao et al. (2021) investigated the personal name usage in children's literature, and stories written by over 100,000 children aged 5-13 in the U.K. Results

showed that male authors overrepresent male names in children's books: Male authors produced 159,579 names, of which 71% (112,998) were male names. The other finding was that boys wrote more about boys with age due to the input they were exposed to: Among 429,804 names produced by boys, 85% (N = 363,189) were male names, and only the remaining 15% (N = 66,615) were female. By contrast, female authors and little girls used more balanced names in their stories. The findings suggest that the written language in children's books can shape and be shaped by gender biases (Hsiao et al., 2021). On a different note, gender equality may be affected by the creeping effect of gendered languages on perception. For instance, Prewitt-Freilino et al. (2011) found that gender equality in countries with a gendered language was significantly worse than in countries with a genderless language.

Another pertinent finding of the study was that the effect of French as an L2 on perception could potentially change the stereotypical gender of the object. For example, the experiment used *car* as a control item because both English and French monolinguals perceived it to be stereotypically associated with masculinity, despite it being feminine in French in the pilot. However, item analyses (to see object-level analyses, please refer to Appendix M) showed that English-dominant bilinguals rated *car* significantly higher (M=51.51) than English monolinguals (M=32.57), ignoring its association with masculinity. In other words, English-dominant bilinguals started to ignore the stereotype and rely more on the grammatical gender after acquiring French as the L2. The same effect was demonstrated with *bird* (masculine in French) and *tower* (feminine in French). The feminine association with *bird* and the masculine association with *tower* were ignored after acquiring the French gender system. These findings might suggest that the effect of a gendered language on perception is stronger than the societal association bias (e.g., *car*'s association with masculinity). Put differently, association bias aside, a gendered language might further impose gender biases. For example, a study found that women tend to be less motivated when filling in questionnaires or surveys written in gendered languages due to the usage of masculine generics (Vainapel et al., 2015). The masculine generics refer to the male-inflected parts of speech used for both sexes in certain gender

languages; for example, *les étudiants* ('the students') in French can refer both to groups of male students and mixed groups of students. As a result, they were establishing gender-fair language strategies (e.g., using paired forms his/her instead of using his to refer to both genders, Lindqvist et al., 2018) could be significant in diminishing gender biases.

5.2. A Genderless L2 Does Not Weaken the Effect of French Grammatical Gender

The third question in this study sought to determine if learning a genderless L2 (English) would impact French speakers' perception of objects. In consonance with predictions, the current study found no significant effect except a marginal difference in descriptive statistics. French-dominant bilinguals (sequential and simultaneous ones) only rated experimental male items slightly lower than French monolinguals but not experimental female objects. Both French-dominant bilinguals and French monolinguals answered the open-ended question primarily by referring to the grammatical gender system. These findings suggest that acquiring a genderless L2 might not effectively reduce the effect of the original L1 grammatical gender system. This notion resonates strongly with previous studies (Bassetti & Nicoladis, 2016; Boroditsky & Schmidt, 2000; Forbes et al., 2008).

Additionally, in this study, most French-dominant bilinguals had an advanced level of English proficiency. It is thus possible to raise the idea that the effect of the French gender system would not be affected by even a high level of English proficiency. This outcome is contrary to that of Sato et al. (2013), who found the effect of the French gender system to be diminished by a high level of English proficiency in French-English bilinguals. Sato and his colleagues recruited 60 French-English sequential adult bilinguals in Switzerland and 61 English-French sequential adult bilinguals in the U.K. A C-test was used to measure L2 proficiency objectively and to assign participants to the advanced (N = 35 for French speakers; N = 40 for English speakers) and intermediate (N = 25 for French speakers; N = 21 for English speakers) groups. Participants were asked to complete an online sensibility judgment task in English and French. The task involved three sentences. The first sentence introduced a stereotypical role noun (plural form)

which can be female (e.g., social workers), male (e.g., surgeon) and neutral (e.g., musicians). For example, the first sentence can be "The *social workers* were walking through the station" (Sato et al., 2013, p. 797). The following second and third sentences formed a choice for participants to choose whether the role noun in the first sentence was male or female; for example, "At the end of the day, the majority of the men seemed to want to go home / At the end of the day, the majority of the women seemed to want to go home" (Sato et al., 2013, p. 797). English speakers were assumed to choose based on the stereotypes. In contrast, French speakers were expected to show a male-dominant choice pattern because these role nouns were all masculine generics (can refer to both genders but have a masculine-only interpretation simultaneously). Results showed that when the task language was switched from English to French, French-English bilinguals moved from male-dominant choices to stereotypical choices. On the contrary, English-French bilinguals produced more male-dominant answers when the language was switched from English to French. Another finding was that participants who were highly proficient in L2 resembled native speakers of their L2, and less proficient participants were more influenced by their native languages (Sato et al., 2013).

A discrepancy appears in the findings: Sato et al. (2013) found that acquiring genderless L2 English at a high proficiency level did affect the French gender system, while this study found no significant effect. Further, the finding of Sato et al. (2013) was not in line with other evidence in literature either (Bassetti & Nicoladis, 2016). This discrepancy could be due to the different focus of comparison of the study. Sato et al. (2013) compared English-French bilinguals' responses with French-English bilinguals, not French monolinguals, as in this study. Resemblances between English-French bilinguals' responses and French-English bilinguals can hardly lead to the conclusion that acquiring English had an effect. The lack of French monolinguals as a baseline for comparison could be an issue. However, because of the objective C-test, Sato et al. (2013) showed that growth in L2 proficiency can play a role; this is an additional difference in findings compared to this study (discussed in the next section 5.3).

5.3. Independence from L2 Proficiency and Other Demographic Variables

Another important finding was that a higher level of proficiency in a gendered L2 did not indicate a more potent effect of the gender system on perception; further, neither did a higher level of proficiency in a genderless L2 lead to a weaker impact of the L1 grammatical gender system. These findings coincide pertinently with Kurinski and Sera's (2011), where the effect of the Spanish gender system on the perception of L1 English L2 Spanish learners did not increase with their proficiency in Spanish. This study strengthens the analysis of Kurinski and Sera (2011) by adding the L2 proficiency test. Thus, the results further support the idea that for speakers of a genderless L1, the effect of gender on perception might grow independently of the gendered L2 proficiency. A note of caution is due here since literature shows that, for speakers of a gendered L1, a higher proficiency in another gendered L2 would weaken the effect of the L1 grammatical gender system (e.g., Bassetti & Nicoladis, 2016). Therefore, the combination of bilingual or multilingual languages is important when interpreting the role of L2 proficiency in the effect of the L1 gender system.

Likewise, for speakers of a gendered L1, the impact of a gendered L1 may potentially not reduce with an increase in genderless L2 proficiency. However, as stated in the previous section 5.2, this finding is contrary to Sato et al. (2013), who found that an increase in English proficiency decreases the effect of the L1 French gender system. The different usage of tasks to measure perception could cause such a difference. This study used a voice distribution task, while Sato et al. (2013) adopted the rarely used sensibility judgment task (Samuel et al., 2019). The voice distribution task has been found as likely to yield supporting results of linguistic relativity, as shown in the systematic review conducted by Samuel et al. (2019). A common criticism of using voice tasks is that its instruction explicitly contains gender. As a result, it might increase the chances of participants unconsciously referring to the grammatical gender system to complete the task (Bender et al., 2011). For example, when justifying why s/he gave a feminine rating to *table*, an English monolingual participant in this study mentioned the French grammatical gender system. Perhaps the consent sheet in both languages and the

instruction of the voice task hinted at the participant. However, Sato et al.'s (2013) sensibility judgment task (see details in the previous section 5.2) also has its limitations. For example, the task was based on stereotypes; the association between *social workers* and femineity is hard to perceive. Also, no pre-test was conducted to see if participants' preconceived connotations about these stereotypes were congruent with the authors' assumption. These limitations for both tasks resonate with a criticism of this line of linguistic relativity research: results depend heavily on the task types (Samuel et al., 2019). There is ample room for further progress in determining the role of genderless L2 proficiency. Adopting a neurolinguistic approach, using different tasks, and proper forms of L2 proficiency testing are helpful.

The proficiency measure turned out to be crucial. A significant discrepancy was found between English participants' self-perceived higher French proficiency and their actual lower capability. On the contrary, French-dominant bilinguals tended to underestimate their English proficiency. These findings consider that previous studies using participants' self-perceived proficiency as the measure for L2 proficiency (e.g., Lambelet, 2016) may lack validity to a certain extent, rendering their results less convincing. A possible explanation for French high proficient English learners underestimating their English proficiency could potentially be foreign language anxiety (a particular combination of perceptions, beliefs, feelings, and behaviours resulting from unique language learning, Horwitz, 1986) (MacIntyre et al., 1997; Hewitt & Stephenson, 2012). Other possible reasons include top performers' tendency to be modest (Ehrlinger et al., 2008). However, English participants with a relatively low level of French displayed a sense of 'overconfidence' in their L2 abilities. Such a contrast between English and French participants resonates with a previous study conducted by Trofimovich et al. (2016), in which high proficient learners underestimated their L2 abilities while less proficient learners overestimated them. The tendency for poorer learners to misjudge their capabilities has been conceptualised and demonstrated as the Dunning-Kruger effect (Dunning et al., 2003). In this sense, English-dominant bilinguals in this study could not calibrate their self-assessment with true abilities because they were potentially unaware of their inabilities.

Apart from L2 proficiency, other demographic variables (metalinguistic awareness, L2 exposure, starting age of L2) were found not to predict the ratings for objects. The self-reported data for metalinguistic awareness and L2 exposure may limit the interpretation (Dörnyei & Taguchi, 2009). Participants could potentially report a range of L2 exposure inaccurate from their accurate language contact. It may also be possible that metalinguistic awareness affects how participants learn the grammatical gender system (e.g., French words ending with consonants are usually masculine) (Brooks & Kempe, 2012), but not how the gender system creeps into perception. Regarding simultaneous and sequential bilinguals, the study found no difference between these two kinds of bilinguals might be because of the small sample size of simultaneous bilinguals. The study only included one simultaneous French-dominant bilingual and nine French-dominant ones. The unequal sample size can limit the rigour of data analysis (Shaw & Mitchell-Olds, 1993).

5.4. French Might Introduce Gender Stereotypes

As previously discussed, French as an L2 seemed to impact the stereotypical gender of particular objects, such as *car*, *tower*, and *bird*. However, with closer inspection, the French might induce gender biases into people's minds. An unexpected finding of the study was that the results of the comparison of four levels all (somewhat) clustered in control, feminine items. Although the comparison of French monolinguals and French-dominant bilinguals showed no significant difference, the descriptive statistics displayed a slight difference in ratings. Control items in this study were carefully chosen based on the pilot study; they had to be the objects associated with typical gender biases. The results of the RQ1 showed that French monolinguals gave a significantly higher gender rating in control feminine objects. Results of RQ2 showed that English-dominant bilinguals also rated control, female objects significantly higher than English monolinguals. These findings could potentially suggest that the acquisition of French as an L2 can induce female gender biases of objects in English participants.

The acquisition of English as an L2 might not help reduce the gender biases from the initial effect of L1 French. RQ3 showed that French-dominant bilinguals rated control feminine items slightly lower than French monolinguals. Also, in the results of RQ4, French-dominant bilinguals still rated control female objects higher than English-dominant bilinguals. Thus, French might tend to induce gender biases in its speakers and learners. Additionally, such an effect from French appears to remain unaffected by acquiring genderless English. These findings raise intriguing questions regarding the practical implications of debates on gender equality. As noted in their review, Bassetti and Nicoladis (2016) claimed that an essential line of future research on linguistic relativity and grammatical gender is to see how bilingualism impacts gender attitudes. Future research can be undertaken with more language pairs to see how the gender attitude changes with different pairs (one gendered language or two gendered languages, Bassetti & Nicoladis, 2016). Future research can also focus on what kind of gender biases would be introduced by French, as only feminine gender biases were shown in this study (none of the four groups differed significantly in ratings for the control male items).

5.5. Pedagogical Implications

In terms of pedagogical implications, teachers might need to be aware of the effect of the grammatical gender system when teaching an L2 with a gender system. Students could perceive objects as masculine or feminine depending on their L1/L2 and use the wrong grammatical gender in production. Teachers can thus emphasise the discrepancies between perception and grammar during teaching. Emphasising the difference could be important because studies have shown that speakers with a genderless L1 often find the L2 grammatical gender system challenging to acquire (e.g., Sabourin et al., 2006; Shimanskaya & Slabakova, 2019). Also, when speakers perceive an object based on their L1 that is different from its true grammatical gender in the L2, cross-linguistic influence (CLI) might come into play. CLI is the interrelationship (or bi-directional influence) between L1 and L2 that results in the co-activation of the two language systems of a bilingual (Hervé & Serratrice, 2018; Nicoladis et al., 2021). Nicoladis and colleagues (2021) touched on this consideration by investigating

whether the subconscious intuitions that native English speakers have about objects' gender could impact English learners of French in judging the French object's gender. Results showed that participants were more accurate with French words with congruent English gender connotations. This finding indicates the cross-linguistic influence of English on the acquisition of the French gender system (Nicoladis et al., 2021). Therefore, it is suggested that French teachers mention the usual discrepancies between stereotypes and French grammar that English students might have. This technique can facilitate the teaching and learning of pronouns and agreement (the change of forms in adjectives, verbs, articles, etc., that goes with the grammatical gender) in gendered languages. A pertinent limitation of Nicoladis et al. (2021) and this study could be that the sole focus was on English and French. Future research can focus on more combinations of languages, for example, German, as the general literature reflects inconclusive results (as discussed in the literature review). Future research can also focus on the gender system learning of speakers with a non-Indo-European L1 (e.g., Chinese or Japanese). CLI might play a role in language combinations of Indo-European languages because of close morphosyntactic and phonological cues. However, Chinese or Japanese speakers might differ from English speakers in learning French or other gendered languages; the difference might be in the connotation of objects' genders because of cultural differences (Xia & Wang, 2017).

5.6. Limitation and Future Directions

The study tried to answer RQs by utilising rigorous and sound methods appropriately applied to the research questions. However, the results should still be interpreted with caution due to limitations. Based on limitations, future research suggestions are given.

5.6.1. Limitations in Research Design

Firstly, although LexTALE was an objective L2 proficiency test, it can only offer an indication of the vocabulary aspect of L2 proficiency (de Bruin, 2019). A more comprehensive battery of L2 proficiency measurements can be utilised. Secondly, the voice distribution task was chosen

to measure participants' perceptions. However, the task cannot fully reflect people's perceptions. Also, according to the systematic review of Samuel et al. (2019), the voice attribution task tended to yield more supportive evidence for the effect of the grammatical gender system on perception. In contrast, the properties judgment task (in which participants were asked to describe the properties or characteristics) tended to have more disapproving evidence against the same effect. Although the latter properties judgment task might be partially biased because it usually contains gender stereotypes (properties like 'tough' associated with masculinity and 'soft' with femineity), future studies can utilise both voice distribution and properties judgment tasks to add more rigour. Whether participants would give the same answer patterns in these different methods can also be an intriguing direction. Furthermore, the neurolinguistic approach might be a more direct way to measure perception (Athanasopoulos & Casaponsa, 2020).

Further, as the primary recruitment method was through Prolific.com, the participants were from worldwide. However, English French bilingual places like Canada might display unique patterns in their object perception, which can be a potential research question for future research. The study also suffered from self-reported bias in the background information questionnaire; for example, the L2 exposure was a form of self-reported data. Moreover, previous studies noted that participants' perceptions might be influenced by whether the objects were animate things or artefacts (e.g., Bassetti, 2007). Thus, future research can choose objects more selectively by balancing the number of (non-)artefacts or (in-)animate items.

An unavoidable limitation of this study is that only adults were recruited. However, as discussed in the literature review, there was a potential age boundary (seven or eight), after which the language's effect started to play a role (Flaherty, 2001). Therefore, future research would be beneficial to investigate when these effects of the gender system on perception begin to emerge. Such a direction has important theoretical implications regarding the debate on

linguistic determinism and universalism (discussed in the literature review). Neither of the two strong accounts could hold if language influences thought after a certain age.

A degree of participants noticed the study's aim because the consent form was in English and French. To hide the purpose of the study, future research can distract participants to a certain extent. For example, in the instructions, participants can be told that they help to decide on objects' characteristics to create a cartoon (Lambelet, 2016). As for the drawings, one participant wrote that s/he gave more feminine results because the drawing style was curly (moderate association with femineity), which might be attributed to the fact that the female researcher created all the drawings. Though it is hard to resolve in empirical research, future researchers might consider including more drawing styles.

Lastly, the study initially wanted to include German to see the effect of different grammatical gender systems. However, adding German required more groups of participants. The number of participants needed to generate enough power exceeded the researcher's recruiting capability. Thus, German was removed from the study. Future research can take this step further to see if the addition of German could affect the L1 French gender system and to see how the addition could affect gender stereotypes (Bassetti & Nicoladis, 2016).

5.6.2. Future of Linguistic Relativity Research.

On a different note, linguistic relativity research could be more related to CLI research. This consideration is suggested because CLI might complement the role of linguistic relativity in affecting the gender system. For example, as mentioned in the literature review, Lambelet (2016) found a tentative result: French learners might assign gender to objects not inconsistently with the actual French gender, but with the French gender. Thus, participants with different L1s might think of the objects' French gender differently depending on L1 and/or L2 influences. The effect of CLI might skew the (interpretation of) the results of linguistic relativity research. Thus, the two phenomena can be combined. The combination might bring

more pedagogical implications for CLI and linguistic relativity research. For example, teachers can be benefited by focusing on the discrepancies between students' misjudged gender of French words (due to CLI) and the actual gender.

6. Conclusion

This quantitative study explored linguistic relativity by investigating if there is an effect of the French grammatical gender system on French speakers' and learners' perceptions of objects and, if so, the kind of effect occurring. To address this aim, the research project set out four RQs to compare three levels: English vs French monolinguals, English monolinguals vs English-dominant bilinguals, French monolinguals vs French-dominant bilinguals, and English-dominant bilinguals vs French-dominant bilinguals. Data were collected through careful piloting and an online experiment; they were analysed using multi-level modelling and multiple *t*-tests. This study has identified an effect of the French grammatical gender system on monolinguals' and bilinguals' perceptions. The research has also shown that the effect of the French language as an L1 remained unaffected by the acquisition of English. The results hold for both early and sequential bilinguals and show no signs of change with the growth in L2 proficiency. Therefore, linguistic relativity is supported by the current findings. Further, one of the more intriguing findings is that the French might introduce gender biases toward particular objects.

The present study adds to the growing body of research on linguistic relativity that indicates the effect of the grammatical gender system on perception (e.g., Samuel et al., 2019). As this research adopted additional rigorous considerations to the design, for example, a sound sample size, a pilot study, control items, L2 proficiency-testing, and advanced data analysis tools, the results contribute to the debate on whether there is an effect of the grammatical gender system. With the objective L2 proficiency measurements, the findings of this study also contribute to the validation of previous arguments that a genderless L2 (even at a high level of proficiency) cannot impact the effect of the gendered L1 on perception (e.g., Bassetti & Nicoladis, 2016). Pedagogically, if language teachers are teaching L2s that have a grammatical gender system, they can emphasise the discrepancies between students' perceptions and L2 grammar. Highlighting the discrepancies can facilitate teaching gender pronouns or agreement in gendered languages. Furthermore, the findings possibly shed new light on the implications of

bilingualism (with gendered language like French) --- a potential ability to reduce or induce gender stereotypes.

One of the main weaknesses of this study was that the degree of participants was primed by the English-French consent form and might have guessed the aim of the study. Further research might replicate the study with more distraction. The role of animacy of objects can also be included (Flaherty, 2001). Another limitation was to what extent measuring objects' voices could represent participants' perceptions. The voice distribution task also has the potential criticism of yielding supporting evidence to linguistic relativity (Samuel et al., 2019). To improve this issue, future research can adopt different tasks or a neurolinguistic approach to explore linguistic relativity by directly focusing on brain activities (Athanasopoulos & Casaponsa, 2020). Further research on linguistic relativity could also offer significant assistance in linking linguistic relativity and cross-linguistic influence. Such a link might be more helpful in yielding more pedagogical implications for L2 education. Testing different combinations of languages, such as French and German/Chinese, is also recommended to see if the effect of the grammatical gender system increases or decreases with the L2. Studying language pairs other than English and French can also explore the interplay between bilingualism (or multilingualism) and gender stereotypes. Overall, I hope my study will contribute to further research on the grammatical gender system and its effect.

References

Ahearn, L. M. (2017). *Living Language: An Introduction to Linguistic Anthropology* (2nd ed.). Wiley Blackwell.

Antón, E., Carreiras, M., & Duñabeitia, J. A. (2019). The impact of bilingualism on executive functions and working memory in young adults. *PLoS One, 14*(2), 1-30. doi:10.1371/journal.pone.0206770

Athanasopoulos, P. (2009). Cognitive representation of colour in bilinguals: The case of Greek blues. *Bilingualism, 12*(1), 83-95. doi:10.1017/S136672890800388X

Athanasopoulos, P., & Casaponsa, A. (2020). The Whorfian brain: Neuroscientific approaches to linguistic relativity. *Cognitive Neuropsychology, 37*(5-6), 393-412. doi:10.1080/02643294.2020.1769050

Athanasopoulos, P., & Bylund, E. (2020). Whorf in the wild: Naturalistic evidence from human interaction. *Applied Linguistics, 41*(6), 947-970. doi: 10.1093/applin/amz050

Bababekov, Y. J., Hung, Y. C., Hsu, Y. T., Udelsman, B. V., Mueller, J. L., Lin, H. Y., . . . Chang, D. C. (2019). Is the power threshold of 0.8 applicable to surgical science?—Empowering the underpowered study. *The Journal of Surgical Research, 241,* 235-239. doi:10.1016/j.jss.2019.03.062

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1-48. doi:10.18637/jss.v067.i01

Bassetti, B. (2007). Bilingualism and thought: Grammatical gender and concepts of objects in Italian-German bilingual children. *The International Journal of Bilingualism: Cross-Disciplinary, Cross-Linguistic Studies of Language Behaviour, 11*(3), 251-273. doi:10.1177/13670069070110030101

Bassetti, B., & Nicoladis, E. (2016). Research on grammatical gender and thought in early and emergent bilinguals. *The International Journal of Bilingualism: Cross-Disciplinary, Cross-Linguistic Studies of Language Behaviour, 20*(1), 3-16. doi:10.1177/1367006915576824

Bauer, D., McNeish, D., Baldwin, S., & Curran, P. (2020). Analysing nested data: Multilevel modelling and alternative approaches. In A. Wright & M. Hallquist (Eds.), *The Cambridge Handbook of Research Methods in Clinical Psychology* (pp. 426-443). Cambridge University Press. doi:10.1017/9781316995808.039.

Bender, A., Beller, S., & Klauer, K. C. (2011). Grammatical gender in German: A case for linguistic relativity? *Quarterly Journal of Experimental Psychology, 64*(9), 1821-1835. doi:10.1080/17470218.2011.582128

Bialystok, E. (2001). *Bilingualism In Development: Language, Literacy, And Cognition*. Cambridge University Press.

Bialystok, E., Poarch, G., Luo, L., & Craik, F. I. (2014). Effects of bilingualism and aging on executive function and working memory. *Psychology and Aging, 29*(3), 696-705. doi: 10.1037/a0037254

Bickel, B. (2000). Grammar and social practice: On the role of 'culture' in linguistic relativity. *Amsterdam Studies in The Theory and History of Linguistic Science Series 4,* 161-192.

Bidaoui, A. (2017). Revisiting the Arabic diglossic situation and highlighting the socio-cultural factors shaping language use in light of Auer's (2005) model. *International Journal of Society, Culture & Language, 5*(2), 60-72.

Blissett, R. (2017, November 26). *Multilevel Modeling in R.* RPubs by RStudio. https://rpubs.com/rslbliss/r_mlm_ws

Bloomfield, L. (1933). *Language*. H. Holt and company.

Boroditsky, L., & Schmidt, L. A. (2000). Sex, syntax, and semantics. In L. R. Gleitman & A. K. Joshi (Eds.), *Proceedings of the 22nd Conference of the Cognitive Science Society* (pp. 42 – 47). Lawrence Erlbaum Associates.

Bosch, L., & Ramon-Casas, M. (2014). First translation equivalents in bilingual toddlers' expressive vocabulary: Does form similarity matter? *International Journal of Behavioral Development, 38*(4), 317-322. doi: 10.1177/0165025414532559

Boutonnet, B., Athanasopoulos, P., & Thierry, G. (2012). Unconscious effects of grammatical gender during object categorisation. *Brain Research, 1479*, 72-79. doi:10.1016/j.brainres.2012.08.044

Brooks, P. J., & Kempe, V. (2012). Individual differences in adult foreign language learning: The mediating effect of metalinguistic awareness. *Memory & Cognition, 41*(2), 281-296. doi:10.3758/s13421-012-0262-9

Brown, R. W., & Lenneberg, E. H. (1954). A study in language and cognition. *Journal of Abnormal and Social Psychology, 49*(3), 454-462. doi:10.1037/h0057814

Brysbaert, M. (2013). Lextale_FR a fast, free, and efficient test to measure language proficiency in French. *Psychologica Belgica, 53*(1), 23-37. doi:10.5334/pb-53-1-23

Byers-Heinlein, K. (2014). Languages as categories: Reframing the 'one language or two' question in early bilingual development. *Language Learning, 64*(2), 184-201. doi:10.1111/lang.12055

Casasanto, D. (2008). Who's afraid of the big bad Whorf? Crosslinguistic differences in temporal language and thought. *Language learning, 58*(1), 63-79. doi:10.1111/j.1467-9922.2008.00462.x

Chen, Z. (2020). Language's effect on cognitive ability: A glimpse into the conspiracy theories. *Advances in Educational Technology and Psychology, 16*, 215-255.

Clarke, M. A., Losoff, A., McCracken, M. D., & Still, J. (1981). Gender perception in Arabic and English. *Language Learning, 31*(1), 159-169. doi:10.1111/j.1467-1770.1981.tb01377.x

Cohen, J. (1977). *Statistical Power Analysis for The Behavioral Sciences* (Revised ed.). Academic Press.

de Bruin, A. (2019). Not all bilinguals are the same: A call for more detailed assessments and descriptions of bilingual experiences. *Behavioural Science, 9*(3), 33-46. doi:10.3390/bs9030033

de Bruin, A., Treccani, B., & Della Sala, S. (2015). Cognitive advantage in bilingualism: An example of publication bias? *Psychological Science, 26*(1), 99-107. doi:10.1177/0956797614557866

Dunning, D., Johnson, K., Ehrlinger, J., & Kruger, J. (2003). Why people fail to recognize their own incompetence. *Current Directions in Psychological Science: A Journal of The American Psychological Society, 12*(3), 83-87. doi:10.1111/1467-8721.01235

Dörnyei, Z., & Taguchi, T. (2009). *Questionnaires in Second Language Research: Construction, Administration, and Processing* (2nd ed.). Routledge.

Ehrlinger, J., Johnson, K., Banner, M., Dunning, D., & Kruger, J. (2008). Why the unskilled are unaware: Further explorations of (absent) self-insight among the incompetent. *Organizational Behavior and Human Decision Processes, 105*(1), 98-121. doi:10.1016/j.obhdp.2007.05.002

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). GPower 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behaviour Research Methods, 39*(2), 175-191. doi:10.3758/BF03193146

Field, A. P. (2009). *Discovering Statistics Using IBM SPSS Statistics* (3rd ed.). SAGE Publications.

Field, A. P. (2013). *Discovering Statistics Using IBM SPSS Statistics* (4th ed.). SAGE Publications.

Field, A. P. (2018). *Discovering Statistics Using IBM SPSS Statistics* (5th ed.). SAGE Publications.

Firestone, C., & Scholl, B. J. (2016). Cognition does not affect perception: Evaluating the evidence for 'top-down' effects. The Behavioral and brain sciences, 39, e229. doi:10.1017/S0140525X15000965

Flaherty, M. (2001). How a language gender system creeps into perception. *Journal of Cross-Cultural Psychology, 32*(1), 18-31. doi:10.1177/0022022101032001005

Flecken, M., Athanasopoulos, P., Kuipers, J. R., & Thierry, G. (2015). On the road to somewhere: Brain potentials reflect language effects on motion event perception. *Cognition, 141*, 41-51. doi:10.1016/j.cognition.2015.04.006

Flege, J. E., Mackay, I. R. A., & Piske, T. (2002). Assessing bilingual dominance. *Applied Psycholinguistics, 23*(4), 567-598. doi:10.1017/S0142716402004046

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*(1), 68-76.

Frey. (2018). *The SAGE Encyclopedia of Educational Research, Measurement, and Evaluation.* SAGE Publications. doi:10.4135/9781506326139

George, D., & Mallery, P. (2010). *SPSS for Windows Step by Step: A Simple Guide and Reference* (10th ed.). Allyn and Bacon.

Grosjean, F. (1982). *Life with Two Languages: An Introduction to Bilingualism.* Harvard University Press.

Grosjean, F. (1989). Neurolinguists, beware! The bilingual is not two monolinguals in one person. *Brain and Language, 36*(1), 3-15. doi:10.1016/0093-934X(89)90048-5

Hartshorne, J. K., Tenenbaum, J. B., & Pinker, S. (2018). A critical period for second language acquisition: Evidence from 2/3 million English speakers. *Cognition, 177*, 263-277. doi: 10.1016/j.cognition.2018.04.007

Hervé, C., & Serratrice, L. (2018). The development of determiners in the context of French–English bilingualism: A study of cross-linguistic influence. *Journal of Child Language, 45*(3), 767-787. doi:10.1017/S0305000917000459

Hewitt, E., & Stephenson, J. (2012). Foreign language anxiety and oral exam performance: A replication of Phillips's MLJ Study. *The Modern Language Journal, 96*(2), 170-189. doi:10.1111/j.1540-4781.2011.01174.x

Hoffman, L., & Rovine, M. J. (2007). Multilevel models for the experimental psychologist: Foundations and illustrative examples. *Behavior Research Methods, 39*(1), 101-117. doi: 10.3758/BF03192848

Horwitz, E. K. (1986). Preliminary evidence for the reliability and validity of a foreign language anxiety scale. *TESOL Quarterly, 20*(3), 559-562. doi:10.2307/3586302

Hsiao, Y., Banerji, N., & Nation, K. (2021). Boys write about boys: Androcentrism in children's reading experience and its emergence in children's own writing. *Child Development, 92*(6), 2194-2204. doi:10.1111/cdev.13623

Huang, I. C. (2016). 'Everybody learns English, and so do I': The dominance of English and its effects on individuals. *English Today, 32*(1), 28-34. doi:10.1017/S0266078415000504

Keysar, B., Hayakawa, S. L., & An, S. G. (2012). The foreign-language effect: Thinking in a foreign tongue reduces decision biases. *Psychological Science, 23*(6), 661-668. doi: 10.1177/0956797611432178

Konishi, T. (1993). The semantics of grammatical gender: A cross-cultural study. *Journal of Psycholinguistic Research, 22*(5), 519-534. doi:10.1007/BF01068252

Kuo, J. Y. C., & Sera, M. D. (2009). Classifier effects on human categorization: The role of shape classifiers in mandarin Chinese. *Journal of East Asian linguistics, 18*(1), 1-19. doi:10.1007/s10831-008-9036-6

Kupisch, T., Mitrofanova, N., & Westergaard, M. (2021). Phonological vs. natural gender cues in the acquisition of German by simultaneous and sequential bilinguals (German-Russian). *Journal of Child Language, 49*(4), 661-683. doi:10.1017/S0305000921000039

Kurinski, E., & Sera, M. D. (2011). Does learning Spanish grammatical gender change English-speaking adults' categorization of inanimate objects? *Bilingualism, 14*(2), 203-220. doi:10.1017/S1366728910000179

Lakoff, G., & Johnson, M. (1980). *Metaphors We Live By*. University of Chicago Press.

Lambelet, A. (2016). Second grammatical gender system and grammatical gender-linked connotations in adult emergent bilinguals with French as a second language. *The International Journal of Bilingualism: Cross-Disciplinary, Cross-Linguistic Studies of Language Behavior, 20*(1), 62-75. doi:10.1177/1367006915576832

Lemhöfer, K. M., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behaviour Research Methods, 44*(2), 325-343. doi:10.3758/s13428-011-0146-0

Lindqvist, A., Renström, E. A., & Gustafsson Sendén, M. (2018). Reducing a male bias in language? Establishing the efficiency of three different gender-fair language strategies. *Sex Roles, 81*(1-2), 109-117. doi:10.1007/s11199-018-0974-9

MacIntyre, P. D., Noels, K. A., & Clément, R. (1997). Biases in self-ratings of second language proficiency: The role of language anxiety. *Language Learning, 47*(2), 265-287. doi:10.1111/0023-8333.81997008

Matusevych, Y., Alishahi, A., & Backus, A. (2017). The impact of first and second language exposure on learning second language constructions. *Bilingualism, 20*(1), 128-149. doi:10.1017/S1366728915000607

Rogers, J. & Révész, A. (2020). Experimental and quasi-experimental designs. In J. McKinley, & H. Rose (Eds.). *The Routledge Handbook of Research Methods in Applied Linguistics* (pp. 133-143). Routledge.

McLaughlin, B. (1978). *Second-Language Acquisition in Childhood.* Halsted Press division of Wiley.

Miralpeix, I., & Muñoz, C. (2018). Receptive vocabulary size and its relationship to EFL language skills. *International Review of Applied Linguistics in Language Teaching, 56*(1), 1-24. doi:10.1515/iral-2017-0016

Mittelhammer, R., Judge, G. G., & Miller, D. (2000). *Econometric Foundations*. Cambridge University Press.

Montrul, S. (2010). Current issues in heritage language acquisition. *Annual Review of Applied Linguistics, 30*, 3-23. doi:10.1017/S0267190510000103

Nicoladis, E., & Foursha-Stevenson, C. (2012). Language and culture effects on gender classification of objects. *Journal of Cross-Cultural Psychology, 43*(7), 1095-1109. doi:10.1177/0022022111420144

Nicoladis, E., Westbury, C., & Foursha-Stevenson, C. (2021). English speakers' implicit gender concepts influence their processing of French grammatical gender: Evidence for semantically mediated cross-linguistic influence. *Frontiers in Psychology, 12*, 740920. doi:10.3389/fpsyg.2021.740920

Ottenheimer, H. (2013). *The Anthropology of Language: An Introduction to Linguistic Anthropology* (3rd ed.). Wadsworth Cengage Learning.

Oxford Languages. (n.d.). https://languages.oup.com/google-dictionary-en/

Padilla, A. M., & Lindholm-Leary, K. J. (1984). Child bilingualism: The same old issues revisited. In J. L. Martinez (Ed.), *Chicano Psychology* (pp. 369-408), Academic Press.

Pandey, A. (2017, May 3). The language you speak affects how your brain experiences the passage of time. *International Business Times.* https://www.ibtimes.com/language-you-speak-affects-how-your-brain-experiences-passage-time-2533953.

Paradis, J., Genesee, F., & Crago, M. B. (2021). *Dual Language Development & Disorders: A Handbook on Bilingualism and Second Language Learning* (3rd ed.). Paul H. Brookes Publishing Co.

Pavlenko, A. (2003). Eyewitness memory in late bilinguals: Evidence for discursive relativity. *The International Journal of Bilingualism: Cross-Disciplinary, Cross-Linguistic Studies of Language Behavior, 7*(3), 257-281. doi:10.1177/13670069030070030301

Pearson, B. Z., Fernández, S., & Oller, D. K. (1995). Cross-language synonyms in the lexicons of bilingual infants: One language or two? *Journal of Child Language, 22*(2), 345-368. doi:10.1017/S030500090000982X

Perlovsky, L. (2009). Language and emotions: Emotional Sapir–Whorf hypothesis. *Neural Networks, 22*(5), 518-526. doi:10.1016/j.neunet.2009.06.034

Phillips, W., & Boroditsky, L. (2003). Can quirks of grammar affect the way you think? Grammatical gender and object concepts. In R. Alterman & D. Kirsh (Eds.), *Proceedings*

of the 20-fifth Annual Meeting of the Cognitive Science Society (pp. 928-933). Cognitive Science Society.

Pinker, S. (1994). *The Language Instinct*. W. Morrow and Co.

Prewitt-Freilino, J. L., Caswell, T. A., & Laakso, E. K. (2011). The Gendering of language: A comparison of gender equality in countries with gendered, natural gender, and genderless languages. *Sex Roles, 66*(3-4), 268-281. doi:10.1007/s11199-011-0083-5

Reader, T. (2019, June 21). *Fixed and Random Effects with Tom Reader* [Video]. YouTube. https://www.youtube.com/watch?v=FCcVPsq8VcA&t=218s

Romaine, S. (1999). Bilingual language development. In M. Barrett (Ed.), *The Development of Language* (pp. 251-276), Psychology Press.

Sabourin, L., Brien, C., & Burkholder, M. (2014). The effect of age of L2 acquisition on the organization of the bilingual lexicon: Evidence from masked priming. *Bilingualism, 17*(3), 542-555. doi:10.1017/S1366728913000643

Sabourin, L., Stowe, L. A., & de Haan, G. J. (2006). Transfer effects in learning a second language grammatical gender system. *Second Language Research, 22*(1), 1-29. doi:10.1191/0267658306sr259oa

Samuel, S., Cole, G., & Eacott, M. J. (2019). Grammatical gender and linguistic relativity: A systematic review. *Psychonomic Bulletin and Review, 26*(6), 1767-1786. doi:10.3758/s13423-019-01652-3

Sapir, E. (1956). Language. In D. G. Mandelbaum (Ed.), C*ulture, Language and Personality: Selected Essays* (pp. 1-44). University of California Press.

Sato, S., & Athanasopoulos, P. (2018). Grammatical gender affects gender perception: Evidence for the structural - feedback hypothesis. *Cognition, 176*, 220-231. doi:10.1016/j.cognition.2018.03.014

Sato, S., Gygax, P. M., & Gabriel, U. T. E. (2013). Gender inferences: Grammatical features and their impact on the representation of gender in bilinguals. *Bilingualism, 16*(4), 792-807. doi:10.1017/S1366728912000739

Sebastián-Gallés, N., Echeverría, S., & Bosch, L. (2005). The influence of initial exposure on lexical representation: Comparing early and simultaneous bilinguals. *Journal of Memory and Language, 52*(2), 240-255. doi:10.1016/j.jml.2004.11.001

Sera, M. D., Elieff, C., Forbes, J., Clark Burch, M., Rodríguez, W., & Poulin Dubois, D. (2002). When language affects cognition and when it does not: An analysis of grammatical gender and classification. *Journal of Experimental Psychology, 131*(3), 377-397. doi:10.1037/0096-3445.131.3.377

Shaw, R. G., & Mitchell-Olds, T. (1993). ANOVA for unbalanced data: An overview. *Ecology, 74*(6), 1638-1645. doi:10.2307/1939922

Shimanskaya, E., & Slabakova, R. (2019). L1–L2 differences in the L2 classroom: Anticipating Anglophone learners' difficulties with French pronoun interpretation. *Language Teaching Research, 23*(2), 259-277. doi:10.1177/1362168817739650

Tao, J., Shao, Q., & Gao, X. (2017). Ethics-related practices in Internet-based applied linguistics research. *Applied Linguistics Review, 8*(4), 321-353. doi:10.1515/applirev-2016-2024

Thompson, G., & Dooley, K. (2019). Ensuring translation fidelity in multilingual research. In J. McKinley, & H. Rose (Eds.). *The Routledge Handbook of Research Methods in Applied Linguistics* (pp. 63-75). Routledge.

Treffers-Daller, J., & Silva-Corvalán, C. (2016). *Language Dominance in Bilinguals: Issues of Measurement and Operationalization.* Cambridge University Press.

Trofimovich, P., Isaacs, T., Kennedy, S., Saito, K., & Crowther, D. (2016). Flawed self-assessment: Investigating self- and other-perception of second language speech. *Bilingualism, 19*(1), 122-140. doi:10.1017/S1366728914000832

Vainapel, S., Shamir, O. Y., Tenenbaum, Y., & Gilam, G. (2015). The dark side of gendered language: The masculine-generic form as a cause for self-report bias. *Psychological Assessment, 27*(4), 1513-1519. doi:10.1037/pas0000156

Vernich, L., Argus, R., & Kamandulyte-Merfeldiene, L. (2017). Extending research on the influence of grammatical gender on object classification: A cross-linguistic study comparing Estonian, Italian and Lithuanian native speakers. *Eesti Rakenduslingvistika Ühingu Aastaraamat, 13*, 223-240. doi:10.5128/ERYa13.14

Vigliocco, G., Vinson, D. P., Paganelli, F., & Dworzynski, K. (2005). Grammatical gender effects on cognition: Implications for language learning and language use. *Journal of Experimental Psychology, 134*(4), 501-520. doi:10.1037/0096-3445.134.4.501

Wasserman, B. D., & Weseley, A. J. (2009). Qué? Quoi? Do languages with grammatical gender promote sexist attitudes? *Sex Roles, 61*(9-10), 634-643. doi:10.1007/s11199-009-9696-3

Whorf, B. L. (1956). *Language, Thought, and Reality: Selected Writings of Benjamin Lee Whorf.* M.I.T. Press.

Williams, R. A. (2020). *Ordinal Independent Variables*. SAGE Publications Limited.

Xia, H., & Wang, Z. (2017). Sanyu xide zhongde eryu qianyi xianxiang jiqi jiaoxue qishi. [Language transfer from second language to the third language and its teaching inspiration]. *Journal of Xihua University (Philosophy & Social Sciences), 36*(5), 103-107.

Zubin, D. A., & Köpcke, K. M. (1984). Affect classification in the German gender system. *Lingua, 63*(1), 41-96. doi:10.1016/0024-3841(84)90031-7

Appendix A

Consent Form and Information Sheet

UNIVERSITY OF OXFORD

UNIVERSITÉ D'OXFORD
DEPARTMENT OF EDUCATION

DÉPARTEMENT D'ÉDUCATION

15 Norham Gardens, Oxford, OX2 6PY
Te(é)l: (+44)(0)1865274024 Fax: (+44)(0)1865274027
general.enquiries@education.ox.ac.uk

Zhuohan Chen, MSc Student/Étudiante de Master en Science
Student Castle Oxford, Osney Lane, OX1 1TE
(+44)(0)7784606912 (UK/Royaume-Uni), (+86)18072731699 (China/Chine)
zhuohan.chen@education.ox.ac.uk

Exploring Linguistic Relativity: The Effect of the Grammatical Gender System on Bilingual Adults' Perception of Objects

CUREC Approval Reference: CIA-22-100

Explorer la Relativité Linguistique: l'Effet du Système de Genre Grammatical sur la Perception des Objets chez les Adultes Bilingues

Référence d'Approbation CUREC : CIA-22-100

General Information

The aim of this study is to explore the interaction between the language you are using and how you feel about the world.

We appreciate your interest in participating in this experiment. You have been invited to participate as you are 18 years of age or over and have English or French as your first or second language. Please read through this information before agreeing to participate (if you wish to) by ticking the 'yes' box below.

Informations générales

Cette étude a pour but d'explorer l'interaction entre la langue que vous utilisez et votre façon de percevoir le monde.

Nous apprécions votre intérêt à participer à cette enquête. Si vous avez été invité à y participer, c'est parce que vous avez 18 ans ou plus et que vous avez l'anglais ou le français comme première ou deuxième langue. Veuillez lire ces informations avant d'accepter d'y participer (si vous voulez) en cochant la case « Oui » ci-dessous.

You may ask any questions before deciding to take part by contacting the researcher (details below).

The Principal Researcher is Zhuohan Chen, who is attached to the Department of Education at the University of Oxford. This project is being completed under the supervision of Docteur Faidra Faitaki.

Avant de décider d'y participer, vous pouvez poser toutes vos questions en contactant la chercheuse (détails ci-dessous).

La Chercheuse Principale est Zhuohan Chen, étudiante au Département d'Éducation de l'Université d'Oxford. Ce projet est effectué sous la supervision du Docteur Faidra Faitaki.

This experiment consists of 3 parts: a general language background check, a small-scale language proficiency test, and a voice decision task. The voice decision task generally asks you to categorise masculine and feminine voices in a continuum for purely academic purposes. This should take about 10-15 minutes. No background knowledge is required. The data is needed to support the quantitative statistical analysis of the dissertation. All data will be anonymised and will be non-identifiable. The data will be only accessible to the researcher Zhuohan Chen and her supervisor Docteur Faidra Faitaki during the research period and will be presented in the appendix of the final dissertation.

Cette enquête est composée de 3 parties : une vérification des connaissances linguistiques générales, un test simple de compétences linguistiques et une tâche de décision vocale. Généralement, la tâche vous demande de continuellement catégoriser les voix masculines (son grave) et féminines (son aiguë) à des buts purement académiques. Cela prendrait environ 10-15 minutes. Aucune connaissance de base n'est requise. Les données sont requises pour soutenir l'analyse statistique quantitative du mémoire. Toutes les données seront rendues anonymes, au lieu d'être identifiables. Les données ne seront accessibles qu'à la chercheuse Zhuohan Chen et à son superviseur Docteur Faidra Faitaki pendant la période de recherche et seront présentées dans l'annexe du mémoire final.

Do I have to take part?

No. Please note that participation is voluntary. If you do decide to take part, you may withdraw at any point for any reason before submitting your answers by pressing the 'Exit' button/ closing the browser. Upon completion, you will join the competition of winning 3 £50 Amazon vouchers. However, we are only able to reimburse participants who complete all study activities.

We have included a 'Prefer not to say' option for each set of questions should you prefer not to answer a particular question.

Est-ce que je suis obligé d'y participer ?

Non. Il est à noter que la participation est volontaire. Si vous décidez d'y participer, avant de soumettre vos réponses, vous pouvez vous retirer à tout moment et pour n'importe quelle raison en appuyant sur le bouton « Quitter » / en fermant le navigateur. Une fois fini, vous vous joindrez à la compétition de gagner 3 coupons Amazon de 50 £. Cependant, seuls les participants qui terminent toutes les activités de l'enquête seront remboursés.

Nous avons inclus une option « Je ne veux pas le dire » pour chaque question si vous ne voulez pas répondre à une question spécifique.

How will my data be used?

We will not collect any data that could directly identify you.

Your IP address will not be stored. We will take all reasonable measures to ensure that data remain confidential.

The responses you provide will be stored in a password-protected electronic file on University of Oxford secure servers and may be used in academic publications, conference presentations, and reports for external organisations. Research data will be stored for **3** years after publication or public release of the work of the research.

Comment mes données seront-elles utilisées ?

Aucune donnée susceptible de vous identifier directement ne sera collectée.

Votre adresse IP ne sera pas stockée. Toutes les mesures raisonnables seront prises afin d'assurer la confidentialité des données.

Les réponses fournies par vous seront stockées dans un fichier électronique protégé par mot de passe sur les serveurs sécurisés de l'Université d'Oxford et pourront être utilisées dans des publications académiques, des présentations de conférence ainsi des rapports pour les organisations externes. Les données de recherche seront stockées pendant 3 ans après la publication ou la diffusion publique du travail de la recherche.

Who will have access to my data?

The University of Oxford is the data controller with respect to your personal data and, as such, will determine how your personal data is used in the study. The University will process your personal data for the purpose of the research outlined above. Research is a task that we perform in the public interest. Further information about your rights with respect to your personal data is available from https://compliance.admin.ox.ac.uk/individual-rights.

Qui aura accès à mes données ?

En tant que responsable du traitement de vos données personnelles, l'Université d'Oxford déterminera la façon d'utiliser vos données personnelles dans l'étude. L'Université traitera vos données personnelles aux fins de l'étude susmentionnée. La recherche est une tâche que nous effectuons dans l'intérêt public. Pour en savoir plus sur vos droits concernant vos données personnelles, vous pouvez visiter https://compliance.admin.ox.ac.uk/individual-rights.

We would also like your permission to use the data in future studies, and to share data with other researchers (e.g. in online databases). Data will be de-identified before it is shared with other researchers or results are made public.

The results will be written up for an MSc degree.

En outre, nous demandons votre autorisation d'utiliser les données dans de futures études et de partager des données avec d'autres chercheurs (par exemple dans des bases de données en ligne). Les données seront rendues anonymes avant le partage avec d'autres chercheurs ou la publication des résultats.

Les résultats seront rédigés pour un diplôme de Master en Science.

Who has reviewed this study?

This project has been reviewed by, and received ethics clearance through, a subcommittee of the University of Oxford Central University Research Ethics Committee [**CIA-22-100**].

Qui a examiné cette étude ?

Ayant reçu une autorisation éthique, ce projet est examiné par un Sous-comité du Comité d'Éthique de la Recherche de l'Université Centrale d'Oxford [CIA-22-100].

Who do I contact if I have a concern or I wish to complain?

If you have a concern about any aspect of this study, please speak to Zhuohan Chen by email at zhuohan.chen@education.ox.ac.uk, or on (+44)(0)7784606912, or her supervisor Docteur Faidra Faitaki by email at faidra.faitaki@education.ox.ac.uk, or on +44(0)7768236586, and we will do our best to answer your query. I/ We will acknowledge your concern within 10 working days and give you an indication of how it will be dealt with. If you remain unhappy or wish to make a formal complaint, please contact the Chair of the Research Ethics Committee at the University of Oxford who will seek to resolve the matter as soon as possible:

Education Departmental Research Ethics Committee, student.curec@education.ox.ac.uk, Dept of Education, 15 Norham Gardens, Oxford, OX2 6PY

Qui dois-je contacter si j'ai un problème ou si je veux porter plainte ?

En cas de doute concernant n'importe quel aspect de cette étude, veuillez contacter Zhuohan Chen par e-mail à zhuohan.chen@education.ox.ac.uk, ou au (+44)(0)7784606912, ou à son superviseur, Docteur Faidra Faitaki par e-mail à faidra.faitaki@education.ox.ac.uk, ou au +44(0)7768236586. Nous nous essayerons de

répondre à votre demande. Nous confirmerons votre doute dans les 10 jours ouvrables et vous donnerons une indication de la résolution.

Si vous n'êtes toujours pas content ou si vous voulez déposer une plainte officielle, veuillez contacter le Président du Comité d'Éthique de la Recherche de l'Université d'Oxford qui essayera de résoudre le problème dès que possible :

Comité d'Éthique de la Recherche du Département d'Éducation, student.curec@education.ox.ac.uk , Département d'Éducation, 15 Norham Gardens, Oxford, OX2 6PY

Please note that you may only participate in this survey if you are 18 years of age or over.

☐ I certify that I am 18 years of age or over

If you have read the information above and agree to participate with the understanding that the data you submit will be processed accordingly, please tick the box below to start.

☐ Yes, I agree to take part

Il est à noter que vous ne pouvez pas participer à cette enquête sauf si vous avez 18 ans ou plus.

☐ Je déclare que j'ai 18 ans ou plus

Si vous acceptez d'y participer après avoir lu les informations ci-dessus et compris que les données soumises par vous seront traitées en conséquence, veuillez cocher la case ci-dessous pour commencer.

☐ Oui, j'accepte d'y participer

Appendix B

Background Information Questionnaire

In this section, you will answer some general questions about your language background.

Dans cette section, veuillez répondre à quelques questions générales sur vos connaissances linguistiques.

1. What is your mother tongue?

A. English

B. French

C. Others

1. Quelle est votre langue maternelle ?

A. Anglais

B. Français

C. Autres

2. Apart from your mother tongue, do you speak other languages fluently? ()

A. Yes

B. No

2. Outre votre langue maternelle, parlez-vous couramment d'autres langues ?

A. Oui

B. Non

3. If yes, apart from your native language, what language do you speak best ? (If no, please skip this question)

A. English

B. French

C. Others

3. Si oui, en plus de votre langue maternelle, quelle langue parlez-vous le mieux ? (Si non, veuillez sauter cette question)

A. Anglais

B. Français

C. Autres

4. What is the level of your second language?

A. native-like/mother-tongue-like

B. advanced/professional

C. intermediate/fluent

D. beginner

F. Prefer not to say

4. Quel est le niveau de votre deuxième langue ?

A. presque aussi bon que ma langue maternelle

B. avancé/professionnel

C. intermédiaire/courant

D. débutant

E. Je ne veux pas le dire

5. Can you speak any other languages? If so, please specify.

A. No

B. Specify _____

5. Pouvez-vous parler d'autres langues ? Si oui, veuillez préciser.

A. Non

B. Préciser_____

6. What is your percentage of use of your second language every day? Quel est le pourcentage d'utilisation/d'exposition à votre deuxième langue tous les jours ?

A. 0-10%

B. 10-30%

C. 30-50%

D. 50-70%

E. 70-90%

F. >90%

7. At what age did you start learning your second language? À quel âge avez-vous commencé à apprendre votre deuxième langue ?

Please Specify/Veuillez Préciser_____

8. Do you have any language qualifications? Avez-vous des qualifications linguistiques ?

Please Specify/Veuillez Préciser_____

9. What is your gender?

A. Male

B. Female

C. Prefer not to say

9. Quel est votre sexe ?

A. Homme

B. Femme

C. Je ne veux pas le dire

10. What is your age?

A. 18 to 30

B. 31 to 50

C. 50 and above

D. Prefer not to say

10. Quel âge avez-vous ?

A. 18 à 30 ans

B. 31 à 50 ans

C. 50 ans et plus

D. Je ne veux pas le dire

11. What is your education level?

A. Primary school

B. Lower/Higher middle school

C. Vocational school

D. Tertiary education

E. Prefer not to say

11. Quel est votre niveau d'études ?

A. École primaire

B. Collège/Lycée

C. École professionnelle

D. Enseignement supérieur

E. Je ne veux pas le dire

Appendix C

Instruction for Vocabulary Tests and Voice Distribution Task

Vocabulary Tests:

To English participants:

French Vocabulary Test

You will now see a test of French vocabulary which consists of about 80 trials, in each of which you will see a string of letters. Your task is to decide whether this is an existing French word or not. If you think it is an existing French word, you click on "yes", and if you think it is not an existing French word, you click on "no".

If you are sure that the word exists, even though you don't know its exact meaning, you may still respond "yes". But if you are not sure if it is an existing word, you should respond "no".

In this experiment, we use British English rather than American English spelling. For example: "realise" instead of "realize"; "colour" instead of "color", and so on. Please don't let this confuse you. This experiment is not about detecting such subtle spelling differences anyway.

You have as much time as you like for each decision. This part of the experiment will take about 5 minutes. If everything is clear, you can now start.

To French participants:

Test de Vocabulaire Anglais

Vous verrez maintenant un test de vocabulaire anglais. Chaque test se compose d'environ 60 essais, dans chacun desquels vous verrez une chaîne de lettres. Votre tâche est de décider s'il s'agit d'un mot anglais existant ou non. Si vous pensez que c'est un

mot anglais existant, vous cliquez sur « oui », et si vous pensez que ce n'est pas un mot anglais existant, vous cliquez sur « non ».

Si vous êtes sûr que le mot existe, même si vous ne connaissez pas sa signification exacte, vous pouvez toujours répondre « oui ». Mais si vous n'êtes pas sûr qu'il s'agisse d'un mot existant, vous devez répondre « non ».

Dans cette expérience, nous utilisons l'orthographe en anglais britannique plutôt qu'en anglais américain. Par exemple : « realise » au lieu de « realize » ; « colour » au lieu de « color », et ainsi de suite. S'il vous plaît ne laissez pas cela vous confondre. Cette expérience ne vise pas à détecter des différences d'orthographe aussi subtiles de toute façon.

Vous avez autant de temps que vous le souhaitez pour chaque décision. Cette partie de l'expérience prendra environ 5 minutes. Si tout est clair, vous pouvez maintenant commencer.

Voice Distribution Task:

Now, please decide on a voice on the slider to the following object. The task generally asks you to categorise masculine (low-pitch) and feminine (high pitch) voices in a continuum for purely academic purposes. This should take about 5-10 minutes. No background knowledge is required. The data is needed to serve as a pilot for the dissertation.

Maintenant, veuillez sélectionner une voix sur le curseur vers l'objet suivant. Généralement, la tâche vous demande de catégoriser les voix masculines (son grave) et féminines (son aiguë) dans un continuum à des fins purement académiques. Cela prendrait environ 5 -10 minutes. Aucune connaissance de base n'est requise. Les données sont requises pour servir de guide pour le mémoire.

Open-ended Question:

Why do you think the object is masculine/feminine as you rated? Pourquoi pensez-vous que l'objet que vous venez d'évaluer est masculin ou féminin ?

Appendix D

English LexTALE Words List and Answer

0=nonword, 1=word.

0	platery	0
0	denial	1
0	generic	1
1	mensible	0
2	scornful	1
3	stoutly	1
4	ablaze	1
5	kermshaw	0
6	moonlit	1
7	lofty	1
8	hurricane	1
9	flaw	1
10	alberation	0
11	unkempt	1
12	breeding	1
13	festivity	1
14	screech	1
15	savoury	1
16	plaudate	0
17	shin	1
18	fluid	1

19	spaunch	0
20	allied	1
21	slain	1
22	recipient	1
23	exprate	0
24	eloquence	1
25	cleanliness	1
26	dispatch	1
27	rebondicate	0
28	ingenious	1
29	bewitch	1
30	skave	0
31	plaintively	1
32	kilp	0
33	interfate	0
34	hasty	1
35	lengthy	1
36	fray	1
37	crumper	0
38	upkeep	1
39	majestic	1

40	magrity	0
41	nourishment	1
42	abergy	0
43	proom	0
44	turmoil	1
45	carbohydrate	1
46	scholar	1
47	turtle	1
48	fellick	0
49	destription	0
50	cylinder	1
51	censorship	1
52	celestial	1
53	rascal	1
54	purrage	0
55	pulsh	0
56	muddy	1
57	quirty	0
58	pudour	0
59	listless	1
60	wrought	1

Appendix E

French LexTALE Words List and Answer

✓ means the word is an actual word.

Stimulus	Mot?	Stimulus	Mot?	Stimulus	Mot?
cheveux	√	gloque		bouton	√
soumon		lézard	√	capeline	√
cloche	√	sacher		lanière	√
fascine	√	nouer	√	honteur	
huif		occire	√	abêtir	√
semonce	√	écouce		fenêtre	√
canoter	√	osseaux		écureuil	√
infâme	√	rejoute		caddie	√
fourmi	√	escroc	√	détume	
cadenas	√	hache	√	oeuiller	
racaille	√	parchance		balai	√
pourcine		pinceau	√	prioche	
œillet	√	poisson	√	vicelard	√
raplaner		robinet	√	joueux	
plaiser		amadouer	√	agire	
cerveler		peigne	√	éventail	√
endifier		retruire		boutard	
jamain		crayon	√	panier	√
ennemi	√	sentuelle		citrouille	√
pouce	√	alourdir	√	bouilloire	√
metter		marteau	√	parir	
fosse	√	esquif	√	remporter	√
inciter	√	treillage	√	procoreux	
salière	√	dauphin	√	tanin	√
fouet	√	orgueil	√	église	√
cessure		amorce	√	indicible	√
clouer	√	cintre	√	réporce	
mappemonde	√	chameau	√	mignon	√

Appendix F

Object Lists

French Gender	*Condition*	*Objects*	*French Gender*	*Condition*	*Objects*
Feminine (n=16)	Experimental (n=11)	Baguette	Masculine (n=16)	Experimental (n=11)	Bed
		Banana			Fish
		Cliff			Milk
		Cup			Watch
		Fork			Champagne**
		Frog			Cheese**
		House			Cloud**
		Moon			Elephant**
		Church**			Fruit**
		Mountain**			Tree**
		Table**			Sun**
	Control (n=5)	Candle		Control (n=5)	Airplane
		Car			Bike
		Teapot			Bird
		Mouse			Snake
		Tower			Train

** means the objects chosen for eliciting an open-ended answer from participants. These objects were chosen because they were rated significantly differently by English and French monolinguals in the pilot study.

Appendix G

CUREC Approval

15 February 2022 9:32

Dear Zhuohan,

Exploring Linguistic Relativity: The Effect of the Grammatical Gender System on Bilingual Adults' Perception of Objects

Approval number: CIA-22-100

Thank you for sharing the revised information letter/consent form. The above application has been considered on behalf of the Departmental Research Ethics Committee (DREC) in accordance with the procedures laid down by the University for ethical approval of all research involving human participants.

I am pleased to inform you that, on the basis of the information provided to DREC, the proposed research has been judged as meeting appropriate ethical standards, and accordingly, approval has been granted.

Please note that any data collection involving in-person interactions with participants must have an up-to-date COVID-19 fieldwork risk assessment in place. Please refer to the current guidance issued by CUREC during the pandemic, notably COVID-19: CUREC guidance on research involving human participants, https://researchsupport.admin.ox.ac.uk/governance/ethics/coronavirus.

If relevant please also check the CUREC website for their best practice research guides, https://researchsupport.admin.ox.ac.uk/governance/ethics/resources/bpg

Good luck with your research study,

Keep well and safe,

All good wishes,

Katharina (member, DREC)

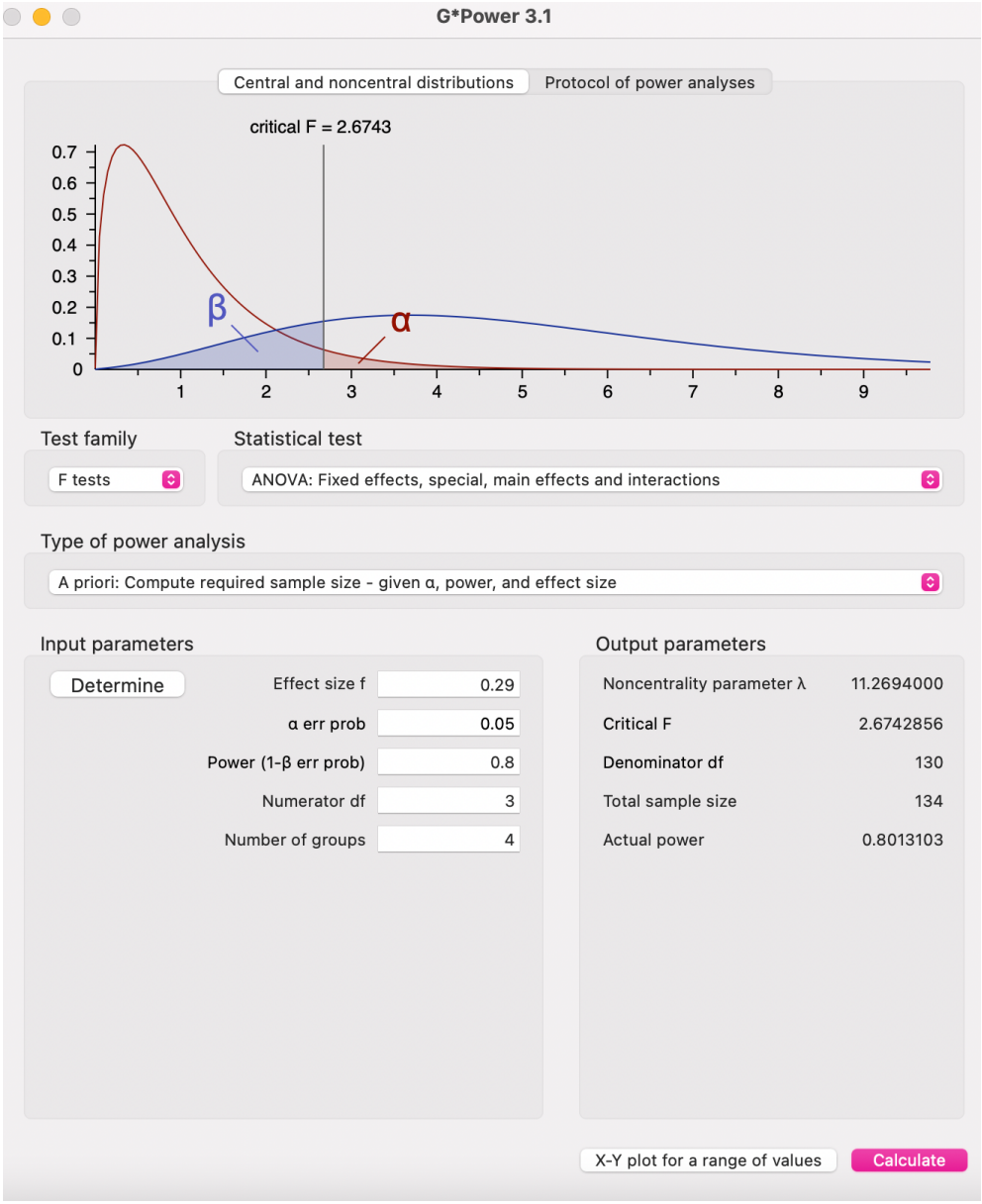
Appendix H

Power Analysis

The study assumed the main statistic method to be ANOVA with main and interaction effects. According to a power analysis (shown in the Figure 2) realised by the G*Power 3 programme, a total of 134 participants (33 per group) would be required to achieve 80% power at the α=.05 level [(1-β)=.8] for main effects. The effect size of .29 was estimated based on Bassetti (2007).

Figure 2

Power Analysis



Appendix I

Cronbach's α

The Cronbach's alpha test was used to test the inter-coder reliability of the codes given by two coders. As seen in Table 19, the codings achieved a high internal consistency, Cronbach's α = .886.

Table 19

Reliability statistics: Cronbach's alpha

Cronbach's Alpha	Cronbach's Alpha Based on Standardised Items	Number of Items
0.886	0.887	2

Appendix J

R Coding for Multi-level Modelling

```
### Statistical analysis using the lme4 package ###
    install.packages("lme4")
    library("lme4")
### Testing the fixed main effects of LanguageGroup, Condition, and ObjectGender,
with random effects for objects and for participants###
    model1 <- lmer(rating ~ Group + (1 |ID) + (1 |Item),
                   data = mydata)
    summary(model1)
    model2 <- lmer(rating ~ condition + (1 |ID) + (1 |Item),
                   data = mydata)
    summary(model2)
    model3 <- lmer(rating ~ gender + (1 |ID) + (1 |Item),
                   data = mydata)
    summary(model3)
### Testing the fixed interaction effects of LanguageGroup*ObjectGender,
LanguageGroup*Condition, and LanguageGroup*Condition*ObjectGender ###
    model4 <- lmer(rating ~ Group*gender + (1 |ID) + (1 |Item),
                   data = mydata)
    summary(model4)
    model5 <- lmer(rating ~ Group*condition + (1 |ID) + (1 |Item),
                   data = mydata)
    summary(model5)
    model6 <- lmer(rating ~ Group*gender*condition + (1 |ID) + (1 |Item),
                   data = mydata)
    summary(model6)
### Comparing models ###
    anova(model1, model2, model3, model4, model5, model6)
```

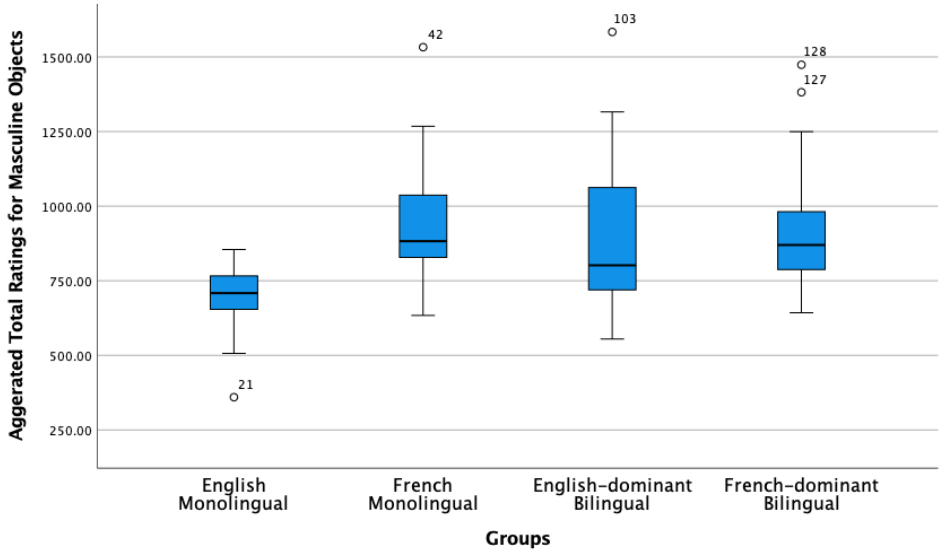
Appendix K

Outliers

As stated in the results chapter, outliers were found using boxplots before the testing of normality. Because the dependent variable (voice ratings) was measured from each participant repeatedly for 32 objects, an aggregated total score would balance out and neutralise the ratings (low ratings represent masculinity and high score represent femineity). Thus, outliers were found in the dependent variable separated by the two within-subject variables (ObjectGender and Condition). All outliers were shown in Graph 18 to Graph 21.

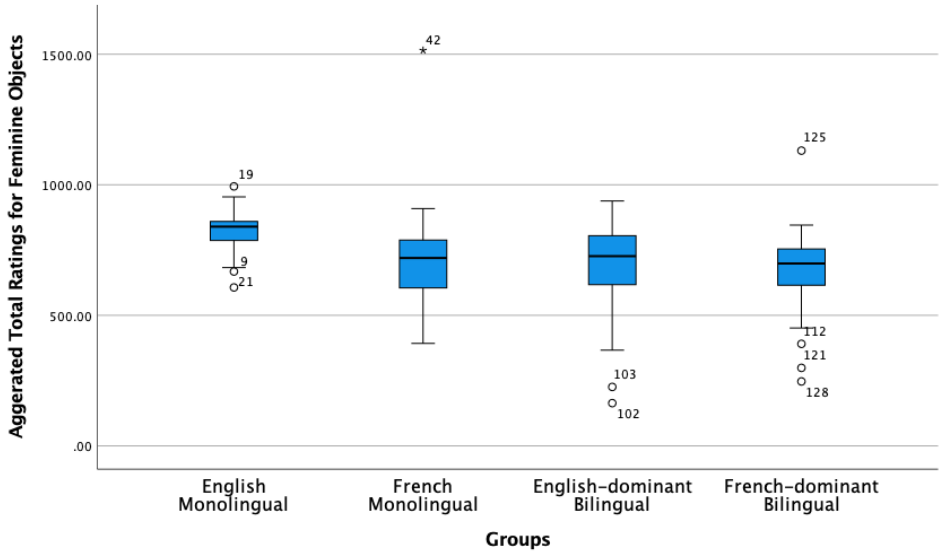
Graph 18

Boxplots for outliers in total ratings for masculine objects of each group



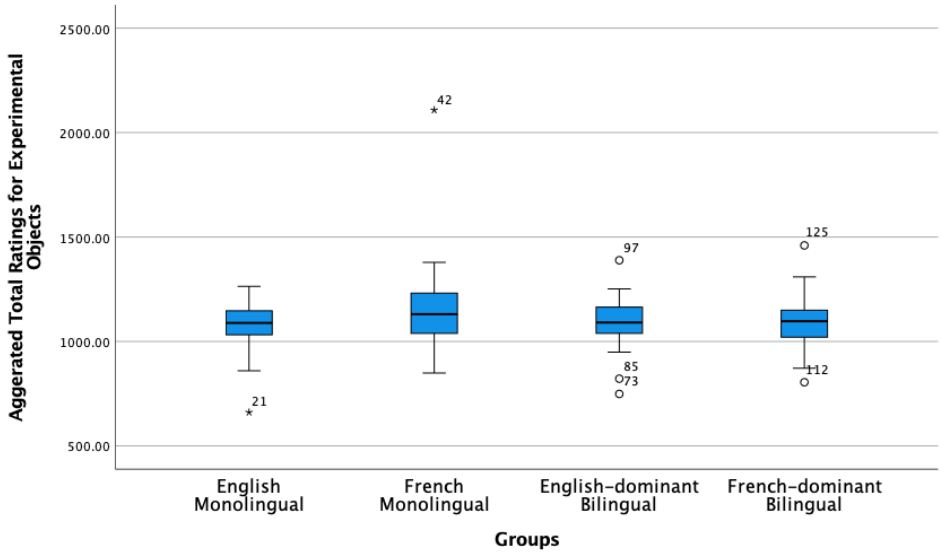
Graph 19

Boxplots for outliers in total ratings for feminine objects of each group



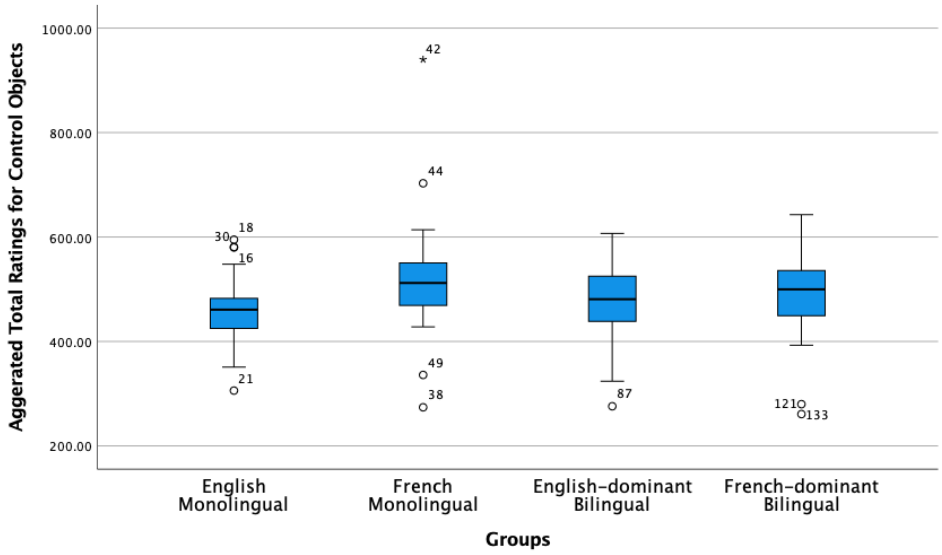
Graph 20

Boxplots for outliers in total ratings for experimental objects of each group



Graph 21

Boxplots for outliers in total ratings for control objects of each group



In all boxplots shown above, the box consists of three horizontal lines, which represent Q1（the first quartile), Q2 (median), and Q3 (the third quartile). The range represents the inter-quartile range. The separate dots in these boxplots are outliers, i.e., the extreme values with a standard deviation larger than |2.8|. In SPSS, the numbers above these dots represent the line number of the extreme data. Therefore, these outliers can be easily located and reassigned to the next highest score (not an outlier) in the group.

Appendix L

Mixed ANOVA Assumptions Unmet

A mixed ANOVA requires the dataset to meet several assumptions: (1) the independent variables are categorical; (2) the dependent variable is continuous and is normally distributed within the groups; (3) for the within-subject variables, the variances of differences at all levels are approximately equal (sphericity); and (4) for between-subject variables, the variance of the dependent variables is approximately equal across groups (homogeneity) (Field, 2018). In this study, the first two assumptions were met. However, SPSS can only tell if the assumption of sphericity and homogeneity are met after the ANOVA is conducted. The following are the main results of the mixed ANOVA.

Table 20

Mixed ANOVA: Within-Subjects Factors and Between-Subject Factors

ObjectGender	Condition	Dependent Variable
Masculine	Experimental	MaleTotal
	Control	FemaleTotal
Feminine	Experimental	Experimental Total
	Control	ControlTotal

		N
Group	English monolingual	35
	French monolingual	35
	English-dominant Bilingual	35
	French-dominant Bilingual	35

Table 21

Mixed ANOVA: Tests of Within-Subjects Effects

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
ObjectGender	Sphericity Assumed	2.445	1	2.445	.001	.980	.000
	Greenhouse-Geisser	2.445	1.000	2.445	.001	.980	.000
	Huynh-Feldt	2.445	1.000	2.445	.001	.980	.000
	Lower-bound	2.445	1.000	2.445	.001	.980	.000
ObjectGender * Group	Sphericity Assumed	1943.091	3	647.697	.175	.913	.004
	Greenhouse-Geisser	1943.091	3.000	647.697	.175	.913	.004
	Huynh-Feldt	1943.091	3.000	647.697	.175	.913	.004
	Lower-bound	1943.091	3.000	647.697	.175	.913	.004
Condition	Sphericity Assumed	18588408.4	1	18588408.4	1061.231	<.001	.886
	Greenhouse-Geisser	18588408.4	1.000	18588408.4	1061.231	<.001	.886
	Huynh-Feldt	18588408.4	1.000	18588408.4	1061.231	<.001	.886
	Lower-bound	18588408.4	1.000	18588408.4	1061.231	<.001	.886
Condition * Group	Sphericity Assumed	588635.891	3	196211.964	11.202	<.001	.198
	Greenhouse-Geisser	588635.891	3.000	196211.964	11.202	<.001	.198
	Huynh-Feldt	588635.891	3.000	196211.964	11.202	<.001	.198
	Lower-bound	588635.891	3.000	196211.964	11.202	<.001	.198
ObjectGender * Condition	Sphericity Assumed	8692345.29	1	8692345.29	545.132	<.001	.800
	Greenhouse-Geisser	8692345.29	1.000	8692345.29	545.132	<.001	.800
	Huynh-Feldt	8692345.29	1.000	8692345.29	545.132	<.001	.800
	Lower-bound	8692345.29	1.000	8692345.29	545.132	<.001	.800
ObjectGender * Condition * Group	Sphericity Assumed	778378.820	3	259459.607	16.272	<.001	.264
	Greenhouse-Geisser	778378.820	3.000	259459.607	16.272	<.001	.264
	Huynh-Feldt	778378.820	3.000	259459.607	16.272	<.001	.264
	Lower-bound	778378.820	3.000	259459.607	16.272	<.001	.264

Table 22

Mixed ANOVA: Levene's Test of Equality of Error Variances

		Levene Statistic	df1	df2	Sig.
Male Total	Based on Mean	11.333	3	136	<.001
	Based on Median	6.640	3	136	<.001
	Based on Median and with adjusted df	6.640	3	94.385	<.001
	Based on trimmed mean	10.898	3	136	<.001
Female Total	Based on Mean	6.768	3	136	<.001
	Based on Median	5.124	3	136	.002
	Based on Median and with adjusted df	5.124	3	98.543	.002
	Based on trimmed mean	6.322	3	136	<.001
Experimental Total	Based on Mean	1.886	3	136	.135
	Based on Median	1.863	3	136	.139
	Based on Median and with adjusted df	1.863	3	122.856	.139
	Based on trimmed mean	1.868	3	136	.138
Control Total	Based on Mean	.867	3	136	.460
	Based on Median	.873	3	136	.457
	Based on Median and with adjusted df	.873	3	128.093	.457
	Based on trimmed mean	.867	3	136	.460

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

As shown in Table 21, the assumption of sphericity for the within-subject variable Condition was not fulfilled ($p < .001$). There were significant differences between the

variances of experimental and control levels. In Table 22, the assumption of homogeneity was also not met. Levene's test showed that regarding ratings for the masculine objects, there were significant differences across all groups, F (3, 136) = 11.333, p <.001. The null hypothesis of equal variances was rejected; in other words, the variances were not equal. Given the other two assumptions of mixed ANOVA were not met, the study could not use it.

Appendix M

Object-level Analysis

One-way ANOVA tests were done for each object, as shown in Table 23. The tests showed if there were any significant differences among the four groups for each specific object. The independent variable was the four groups, which is categorical. The outcome variable was the voice ratings for each object. The assumption of normality was met after reassigning scores to outliers (see earlier). The assumption of homogeneity was shown in Table 24.

Table 23

Object-level Analyses: One-way ANOVA for each object (between groups comparison)

	Sum of Squares	*df*	*Mean Square*	*F*	*Sig.*
Baguette	6884.6	3	2294.9	4.7	0.00
Banana	5805.3	3	1935.1	3.5	0.02
Cliff	8157.5	3	2719.2	4.6	0.01
Cup	2292.6	3	764.2	2.0	0.12
Fork	10432.3	3	3477.4	5.8	<.001
Frog	5245.3	3	1748.4	3.1	0.03
House	7261.8	3	2420.6	4.3	0.01
Moon	8934.3	3	2978.1	5.4	0.00
Church	12884.7	3	4294.9	7.3	<.001
Mountain	8600.7	3	2866.9	4.0	0.01
Table	7769.9	3	2590.0	4.2	0.01
Watch	5699.9	3	1900.0	2.9	0.04
Sun	10491.1	3	3497.0	4.9	0.00
Elephant	7852.9	3	2617.6	5.4	0.00
Fish	1704.0	3	568.0	1.2	0.33
Bed	5883.7	3	1961.2	3.8	0.01
Fruit	2618.7	3	872.9	1.9	0.13
Milk	9850.8	3	3283.6	6.7	<.001
Cloud	9742.3	3	3247.4	5.8	<.001
Champagne	3424.9	3	1141.6	2.2	0.09
Cheese	5854.0	3	1951.3	4.9	0.00
Tree	5025.0	3	1675.0	2.7	0.05
Candle	8963.5	3	2987.8	8.0	<.001
Car	12648.6	3	4216.2	5.9	<.001
Teapot	6506.2	3	2168.7	5.1	0.00
Mouse	4641.7	3	1547.2	3.0	0.03

Tower	3097.5	3	1032.5	2.2	0.10
Bird	6719.2	3	2239.7	3.9	0.01
Airplane	444.4	3	148.1	0.4	0.74
Snake	720.1	3	240.0	0.5	0.66
Bike	1424.3	3	474.8	1.0	0.38
Train	338.3	3	112.8	0.2	0.87

Table 24

Object-level Analyses: Tests of Homogeneity of Variances (Based on Mean)

	Levene Statistic	*df1*	*df2*	*Sig.*
Baguette	2.6	3	136	0.05
Banana	5.6	3	136	0.00
Cliff	3.2	3	136	0.03
Cup	2.6	3	136	0.05
Fork	1.0	3	136	0.39
Frog	1.8	3	136	0.16
House	2.6	3	136	0.05
Moon	1.5	3	136	0.21
Church	1.7	3	136	0.17
Mountain	2.9	3	136	0.04
Table	2.6	3	136	0.06
Watch	2.0	3	136	0.12
Sun	0.3	3	136	0.79
Elephant	0.5	3	136	0.65
Fish	1.6	3	136	0.20
Bed	1.1	3	136	0.37
Fruit	3.5	3	136	0.02
Milk	0.2	3	136	0.87
Cloud	0.4	3	136	0.76
Champagne	1.9	3	136	0.13
Cheese	1.6	3	136	0.19
Tree	0.7	3	136	0.58
Candle	0.5	3	136	0.70
Car	7.4	3	136	<.001
Teapot	1.9	3	136	0.13
Mouse	1.4	3	136	0.26
Tower	2.1	3	136	0.11
Bird	3.9	3	136	0.01
Airplane	0.3	3	136	0.81
Snake	0.6	3	136	0.62
Bike	3.4	3	136	0.02

Train	1.8	3	136	0.15

From Table 23 and Table 24, it was found that there were significant differences in the voice ratings of the four groups in object *baguette, banana, cliff, fork, frog, house, moon, church, mountain, table, watch, sun, elephant, bed, fruit, milk, cloud, cheese, tree, candle, car, teapot, mouse*, and *bird*. It is noteworthy that the assumption of homogeneity for object *banana, cliff, mountain, fruit, car, bird*, and *bike* was not met based on mean. However, because the four groups had equal sample sizes (N=35 per group), the F-test was robust enough to trust (Field, 2013). To further see between which two groups the significant difference lies, planned contrasts were conducted, as shown in Table 25 and Table 26.

Table 25

Object-level Analyses: Contrast Coefficients

Contrast	*Group*			
	English Monolingual	French Monolingual	English-dominant Bilingual	French-dominant Bilingual
1	-1	1	0	0
2	-1	0	1	0
3	-1	0	0	1
4	0	-1	1	0
5	0	-1	0	1
6	0	0	-1	1

Table 26

Object-level Analyses: Planned Contrast for One-way ANOVA of Each Object

	Contrast	*Value of Contrast*	*S.E.*	*t*	*df*	*Sig. (2-tailed)*	*95% Confidence Interval*	
							Lower	*Upper*
Baguette	1	7.1	5.3	1.3	136.0	0.2	-3.4	17.6
	2	16.5	5.3	3.1	136.0	0.0	6.0	27.0
	3	-1.0	5.3	-0.2	136.0	0.8	-11.5	9.5
	4	9.4	5.3	1.8	136.0	0.1	-1.1	19.9
	5	-8.1	5.3	-1.5	136.0	0.1	-18.6	2.4
	6	-17.5	5.3	-3.3	136.0	0.0	-28.0	-7.1
Banana	1	17.1	5.0	3.4	57.5	0.0	7.1	27.0

	2	14.0	5.3	2.6	54.6	0.0	3.4	24.6
	3	11.1	4.9	2.3	57.8	0.0	1.2	21.0
	4	-3.1	6.2	-0.5	67.5	0.6	-15.5	9.4
	5	-5.9	5.9	-1.0	68.0	0.3	-17.7	5.9
	6	-2.9	6.2	-0.5	67.4	0.6	-15.3	9.5
Cliff	1	15.9	5.6	2.8	54.4	0.0	4.7	27.1
	2	5.3	5.4	1.0	56.0	0.3	-5.5	16.1
	3	18.7	4.9	3.8	60.8	<.001	9.0	28.5
	4	-10.6	6.7	-1.6	67.8	0.1	-24.0	2.7
	5	2.8	6.3	0.5	65.7	0.7	-9.7	15.3
	6	13.5	6.1	2.2	66.6	0.0	1.3	25.6
Fork	1	19.4	5.9	3.3	136.0	0.0	7.8	31.0
	2	16.3	5.9	2.8	136.0	0.0	4.7	27.9
	3	22.3	5.9	3.8	136.0	<.001	10.7	33.9
	4	-3.1	5.9	-0.5	136.0	0.6	-14.7	8.5
	5	2.9	5.9	0.5	136.0	0.6	-8.7	14.5
	6	6.0	5.9	1.0	136.0	0.3	-5.6	17.6
Frog	1	11.3	5.7	2.0	136.0	0.0	0.1	22.6
	2	8.4	5.7	1.5	136.0	0.1	-2.9	19.7
	3	16.9	5.7	3.0	136.0	0.0	5.7	28.2
	4	-3.0	5.7	-0.5	136.0	0.6	-14.3	8.3
	5	5.6	5.7	1.0	136.0	0.3	-5.7	16.9
	6	8.6	5.7	1.5	136.0	0.1	-2.7	19.9
House	1	14.9	5.7	2.6	136.0	0.0	3.7	26.0
	2	19.2	5.7	3.4	136.0	<.001	8.1	30.4
	3	9.0	5.7	1.6	136.0	0.1	-2.2	20.2
	4	4.4	5.7	0.8	136.0	0.4	-6.8	15.6
	5	-5.9	5.7	-1.0	136.0	0.3	-17.1	5.3
	6	-10.3	5.7	-1.8	136.0	0.1	-21.4	0.9
Moon	1	21.4	5.6	3.8	136.0	<.001	10.3	32.4
	2	14.3	5.6	2.6	136.0	0.0	3.2	25.3
	3	16.8	5.6	3.0	136.0	0.0	5.8	27.9
	4	-7.1	5.6	-1.3	136.0	0.2	-18.2	3.9
	5	-4.5	5.6	-0.8	136.0	0.4	-15.6	6.5
	6	2.6	5.6	0.5	136.0	0.6	-8.5	13.6
Church	1	23.8	5.8	4.1	136.0	<.001	12.4	35.3
	2	21.7	5.8	3.8	136.0	<.001	10.3	33.2
	3	20.4	5.8	3.5	136.0	<.001	8.9	31.8
	4	-2.1	5.8	-0.4	136.0	0.7	-13.6	9.3
	5	-3.5	5.8	-0.6	136.0	0.6	-14.9	8.0
	6	-1.3	5.8	-0.2	136.0	0.8	-12.8	10.1
Mountain	1	17.5	5.9	3.0	62.0	0.0	5.7	29.3

	2	16.7	5.8	2.9	62.7	0.0	5.1	28.4
	3	19.6	6.1	3.2	60.4	0.0	7.4	31.8
	4	-0.8	6.7	-0.1	68.0	0.9	-14.1	12.6
	5	2.1	6.9	0.3	67.8	0.8	-11.8	15.9
	6	2.8	6.9	0.4	67.7	0.7	-10.9	16.5
Table	1	9.8	5.9	1.7	136.0	0.1	-1.9	21.6
	2	12.3	5.9	2.1	136.0	0.0	0.5	24.0
	3	20.9	5.9	3.5	136.0	<.001	9.2	32.7
	4	2.4	5.9	0.4	136.0	0.7	-9.3	14.2
	5	11.1	5.9	1.9	136.0	0.1	-0.7	22.8
	6	8.7	5.9	1.5	136.0	0.1	-3.1	20.4
Watch	1	15.7	6.2	2.6	136.0	0.0	3.5	27.9
	2	7.5	6.2	1.2	136.0	0.2	-4.7	19.7
	3	15.0	6.2	2.4	136.0	0.0	2.8	27.2
	4	-8.2	6.2	-1.3	136.0	0.2	-20.4	4.0
	5	-0.7	6.2	-0.1	136.0	0.9	-12.9	11.4
	6	7.5	6.2	1.2	136.0	0.2	-4.7	19.7
Sun	1	-19.9	6.4	-3.1	136.0	0.0	-32.5	-7.3
	2	-18.0	6.4	-2.8	136.0	0.0	-30.6	-5.4
	3	-21.5	6.4	-3.4	136.0	0.0	-34.1	-8.8
	4	1.9	6.4	0.3	136.0	0.8	-10.7	14.5
	5	-1.5	6.4	-0.2	136.0	0.8	-14.2	11.1
	6	-3.5	6.4	-0.5	136.0	0.6	-16.1	9.2
Elephant	1	-18.9	5.2	-3.6	136.0	<.001	-29.3	-8.5
	2	-12.9	5.2	-2.5	136.0	0.0	-23.3	-2.5
	3	-17.7	5.2	-3.4	136.0	<.001	-28.1	-7.3
	4	6.0	5.2	1.1	136.0	0.3	-4.3	16.4
	5	1.2	5.2	0.2	136.0	0.8	-9.1	11.6
	6	-4.8	5.2	-0.9	136.0	0.4	-15.2	5.6
Bed	1	-7.3	5.4	-1.3	136.0	0.2	-18.0	3.4
	2	-13.4	5.4	-2.5	136.0	0.0	-24.1	-2.6
	3	-17.1	5.4	-3.2	136.0	0.0	-27.8	-6.4
	4	-6.1	5.4	-1.1	136.0	0.3	-16.8	4.6
	5	-9.8	5.4	-1.8	136.0	0.1	-20.6	0.9
	6	-3.7	5.4	-0.7	136.0	0.5	-14.5	7.0
Fruit	1	-10.1	4.5	-2.3	54.7	0.0	-19.1	-1.1
	2	-9.7	4.9	-2.0	51.5	0.1	-19.5	0.0
	3	-10.1	4.3	-2.4	56.7	0.0	-18.7	-1.5
	4	0.4	5.8	0.1	67.3	0.9	-11.2	12.0
	5	0.0	5.3	0.0	67.8	1.0	-10.6	10.6
	6	-0.4	5.6	-0.1	66.3	0.9	-11.7	10.9
Milk	1	-4.2	5.3	-0.8	136.0	0.4	-14.6	6.2

	2	-21.1	5.3	-4.0	136.0	<.001	-31.5	-10.6
	3	-15.1	5.3	-2.9	136.0	0.0	-25.5	-4.6
	4	-16.9	5.3	-3.2	136.0	0.0	-27.3	-6.4
	5	-10.9	5.3	-2.1	136.0	0.0	-21.3	-0.4
	6	6.0	5.3	1.1	136.0	0.3	-4.4	16.4
Cloud	1	-13.8	5.7	-2.4	136.0	0.0	-25.0	-2.6
	2	-17.6	5.7	-3.1	136.0	0.0	-28.8	-6.4
	3	-22.4	5.7	-4.0	136.0	<.001	-33.6	-11.2
	4	-3.8	5.7	-0.7	136.0	0.5	-15.0	7.4
	5	-8.6	5.7	-1.5	136.0	0.1	-19.8	2.6
	6	-4.8	5.7	-0.8	136.0	0.4	-16.0	6.4
Cheese	1	-10.8	4.8	-2.3	136.0	0.0	-20.3	-1.4
	2	-15.1	4.8	-3.2	136.0	0.0	-24.5	-5.6
	3	-16.5	4.8	-3.5	136.0	<.001	-26.0	-7.1
	4	-4.2	4.8	-0.9	136.0	0.4	-13.7	5.2
	5	-5.7	4.8	-1.2	136.0	0.2	-15.1	3.8
	6	-1.5	4.8	-0.3	136.0	0.8	-10.9	8.0
Tree	1	-8.1	5.9	-1.4	136.0	0.2	-19.8	3.7
	2	-15.5	5.9	-2.6	136.0	0.0	-27.3	-3.8
	3	-13.4	5.9	-2.3	136.0	0.0	-25.1	-1.7
	4	-7.5	5.9	-1.3	136.0	0.2	-19.2	4.2
	5	-5.3	5.9	-0.9	136.0	0.4	-17.0	6.4
	6	2.2	5.9	0.4	136.0	0.7	-9.6	13.9
Candle	1	17.0	4.6	3.7	136.0	<.001	7.9	26.2
	2	-0.8	4.6	-0.2	136.0	0.9	-10.0	8.3
	3	13.8	4.6	3.0	136.0	0.0	4.7	22.9
	4	-17.9	4.6	-3.9	136.0	<.001	-27.0	-8.7
	5	-3.2	4.6	-0.7	136.0	0.5	-12.4	5.9
	6	14.6	4.6	3.2	136.0	0.0	5.5	23.8
Car	1	19.0	5.9	3.2	52.2	0.0	7.2	30.8
	2	24.7	5.5	4.5	55.1	<.001	13.7	35.7
	3	8.7	5.7	1.5	53.2	0.1	-2.9	20.2
	4	5.7	7.0	0.8	67.4	0.4	-8.3	19.7
	5	-10.4	7.2	-1.4	67.9	0.2	-24.8	4.0
	6	-16.1	6.9	-2.3	67.8	0.0	-29.8	-2.3
Teapot	1	16.2	4.9	3.3	136.0	0.0	6.4	26.0
	2	3.9	4.9	0.8	136.0	0.4	-5.9	13.6
	3	14.2	4.9	2.9	136.0	0.0	4.5	24.0
	4	-12.3	4.9	-2.5	136.0	0.0	-22.1	-2.6
	5	-2.0	4.9	-0.4	136.0	0.7	-11.8	7.8
	6	10.4	4.9	2.1	136.0	0.0	0.6	20.2
Mouse	1	10.8	5.4	2.0	136.0	0.0	0.0	21.5

	2	-5.2	5.4	-1.0	136.0	0.3	-15.9	5.6
	3	1.2	5.4	0.2	136.0	0.8	-9.5	12.0
	4	-15.9	5.4	-2.9	136.0	0.0	-26.7	-5.2
	5	-9.5	5.4	-1.8	136.0	0.1	-20.3	1.2
	6	6.4	5.4	1.2	136.0	0.2	-4.3	17.1
Bird	1	-6.3	4.8	-1.3	64.2	0.2	-15.8	3.3
	2	-19.0	5.7	-3.4	56.0	0.0	-30.4	-7.7
	3	-10.7	5.3	-2.0	59.1	0.0	-21.4	-0.1
	4	-12.7	6.1	-2.1	64.1	0.0	-25.0	-0.5
	5	-4.5	5.8	-0.8	66.3	0.4	-16.1	7.2
	6	8.3	6.6	1.3	67.5	0.2	-4.8	21.4

Table 25 displayed the pairs of contrast that are being compared. For example, contrast 1 represented the comparison of English and French monolingual groups. Table 26 displayed which contrast showed the significant differences. Concerning the Discussion chapter, for example, English monolinguals and English-dominant bilinguals differed significantly in their ratings for the object *bird.* This piece of finding can be read from contrast 2 for the object bird in Table 26, $p<.05$.

Appendix N

Research Flyers

Appendix O

Drawings

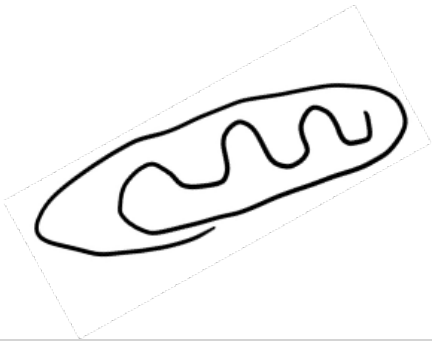
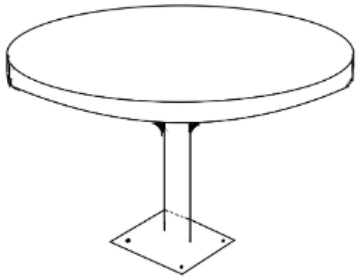
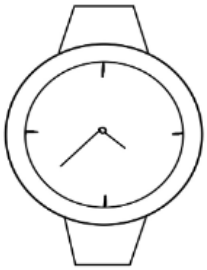
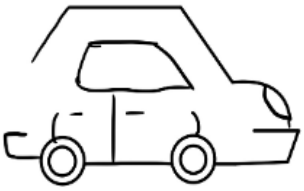
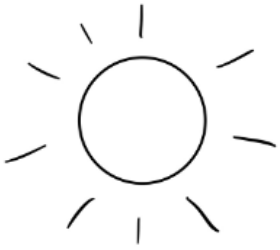
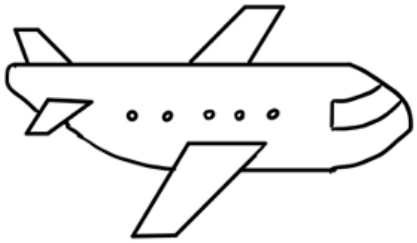
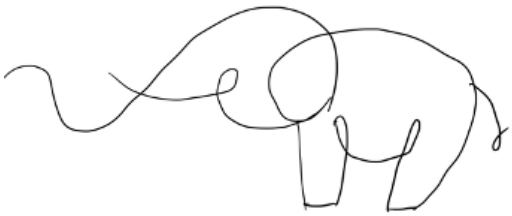
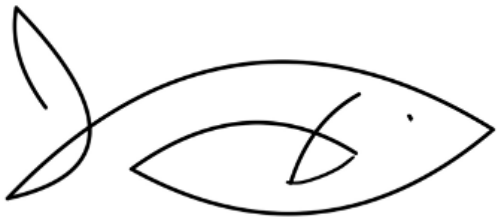
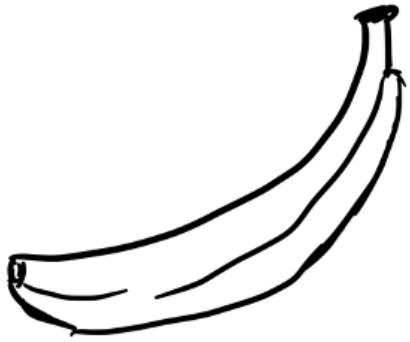
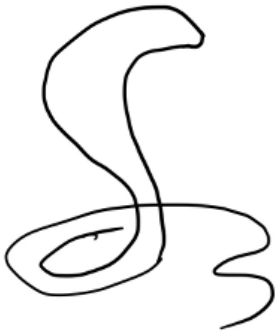
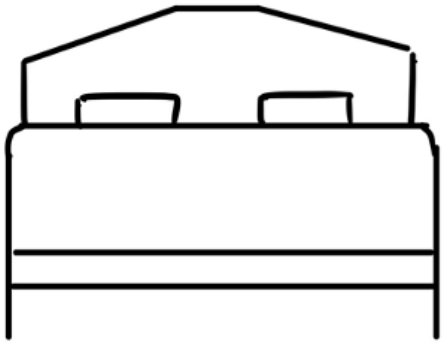
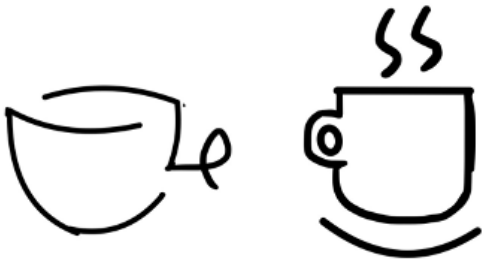
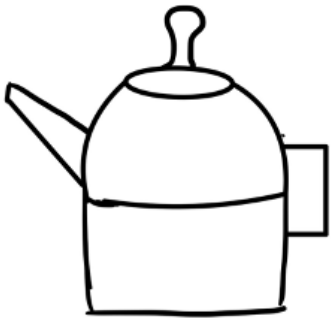
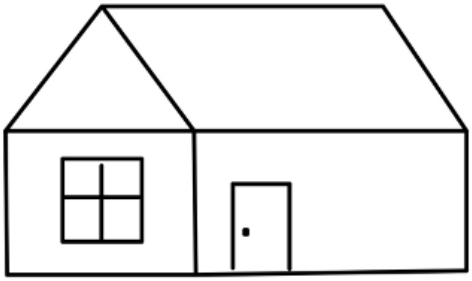
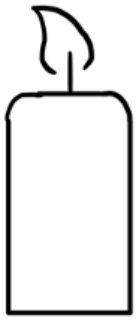
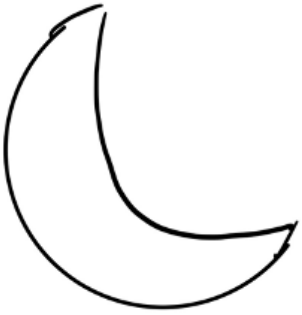
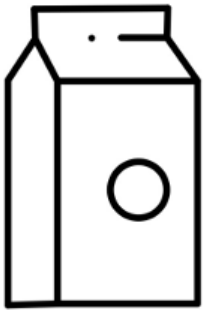
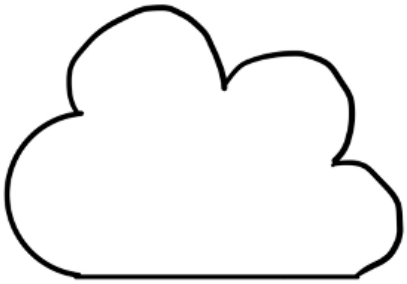
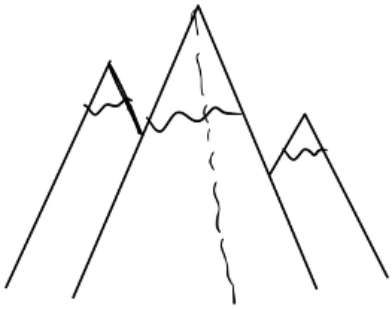
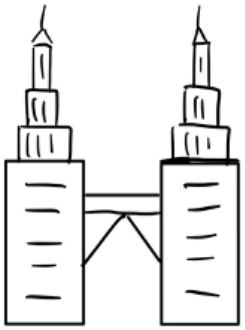
Train	
Baguette	

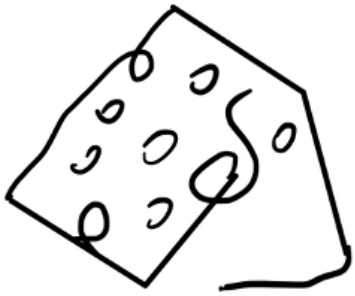
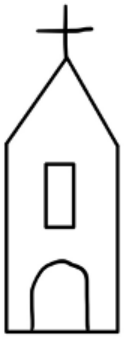
Table	
Watch	
Bird	
Car	
Sun	

Airplane	
Fork	
Elephant	
Fish	
Cliff	

Banana	
Snake	
Bed	
Cup	
Teapot	

Frog	
House	
Candle	
Moon	
Fruit	

Milk	
Cloud	
Mountain	
Tower	
Champagne	

Bike	
Mouse	
Cheese	
Church	
Tree