

HARVARD UNIVERSITY
Graduate School of Arts and Sciences



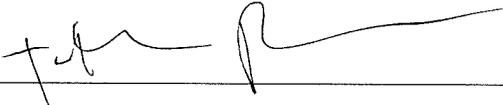
DISSERTATION ACCEPTANCE CERTIFICATE

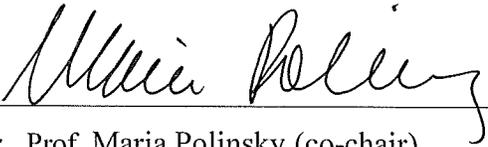
The undersigned, appointed by the
Department of Linguistics
have examined a dissertation entitled

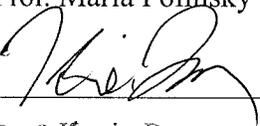
“Gender in the nominal domain: Evidence from bilingualism and eye-tracking”

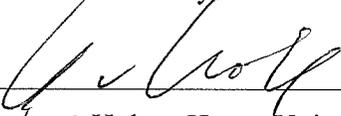
presented by Zuzanna Fuchs

candidate for the degree of Doctor of Philosophy and hereby
certify that it is worthy of acceptance.

Signature 
Typed name: Prof. Jonathan Bobaljik (co-chair)

Signature 
Typed name: Prof. Maria Polinsky (co-chair)

Signature 
Typed name: Prof. Kevin Ryan

Signature 
Typed name: Prof. Holger Hopp, University of Braunschweig

Date: May 3, 2019

Gender in the nominal domain: Evidence from bilingualism and eye-tracking

A DISSERTATION PRESENTED
BY
ZUZANNA FUCHS
TO
THE DEPARTMENT OF LINGUISTICS

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
IN THE SUBJECT OF
LINGUISTICS

HARVARD UNIVERSITY
CAMBRIDGE, MASSACHUSETTS
MAY 2019

©2019 – ZUZANNA FUCHS
ALL RIGHTS RESERVED.

Gender in the nominal domain: Evidence from bilingualism and eye-tracking

ABSTRACT

This dissertation presents two eye-tracking studies, with the goal of investigating heritage speakers' ability to use gender agreement within the nominal domain to facilitate lexical retrieval. Heritage speakers' non-target-like production and comprehension of grammatical gender is well documented, and given these findings, the studies discussed here seek to determine whether and how heritage speakers can deploy gender as it is instantiated in their grammar.

In a Visual World Paradigm context, both control and heritage speakers of Spanish were both able to fixate on target items faster in mismatch conditions, in which the gender on a prenominal definite article was sufficient to disambiguate between the candidate items. The results for the controls replicate the findings of Lew-Williams & Fernald (2007, 2010) and Grüter et al. (2012), while the heritage results provide novel evidence that heritage speakers deploy gender information in online tasks in a monolingual-like manner. Control and heritage speakers of Polish performed the same task, but with gender inflection on a prenominal adjective serving as the critical cue. Again, heritage speakers matched control speakers in their ability to use gender to facilitate lexical retrieval. Additionally, asymmetrical interference between masculine and neuter gender suggests a hierarchical analysis of Polish gender. The Polish heritage speakers matched the controls not only in manner of using gender but also in speed; contrasted with a slight slowdown of the Spanish heritage speakers as compared to the Spanish controls, these findings motivate further study of the difference between accessing determiners and modifiers in online processing of agreement in the nominal domain.

Contents

0	INTRODUCTION	I
1	GRAMMATICAL GENDER	6
1.1	What is (grammatical) gender?	7
1.2	Gender crosslinguistically	10
1.3	Internal organization of gender features and the notion of “default” gender	14
1.4	The relationship between gender and number	16
1.5	Wrap-up	19
2	HERITAGE LANGUAGES & LINGUISTICS	21
2.1	Who are heritage speakers?	23
2.2	Why study heritage languages?	30
2.3	Previous work on heritage languages	34
2.4	The significance of task type	41
2.5	Developmental Trajectories	43
2.6	Contribution of this project	47
3	EYE-TRACKING STUDY: HERITAGE SPANISH	48
3.1	Background	49
3.2	Materials and design	73
3.3	Model interpretations	79
3.4	Participants	80
3.5	Procedure	83
3.6	Proficiency	84
3.7	Analysis & results	94
3.8	Discussion	101
3.9	Going Further	106
4	EYE-TRACKING STUDY: HERITAGE POLISH	109
4.1	Background	112
4.2	Materials and Design	124

4.3	Model interpretations	130
4.4	Participants	130
4.5	Procedure	133
4.6	Proficiency	134
4.7	Analysis & Results	141
4.8	Discussion	155
4.9	Conclusions	160
5	DETERMINERS VERSUS MODIFIERS	163
5.1	Null results	164
5.2	Proficiency	167
5.3	Determiners vs. modifiers	169
5.4	Going forward	186
6	CONCLUSION	188
APPENDIX A EXPERIMENTAL STIMULI		193
A.1	Spanish stimuli	194
A.2	Polish stimuli	196
APPENDIX B LEAP-Q		198
B.1	English version for Spanish speakers	199
B.2	English version for Polish speakers	202
B.3	Spanish version	205
B.4	Polish version	208
APPENDIX C SPANISH GENDER-SPECIFIC ANALYSES		211
C.1	Feminine target item	213
C.2	Masculine target item	216
REFERENCES		236

Listing of figures

3.1	Eye-tracking response times, Lew-Williams and Fernald (2007, 2010).	69
3.2	Spanish Match Condition Display.	77
3.3	Spanish Mismatch Condition Display.	77
3.4	Spanish Auditory Stimulus Frame.	79
3.5	Ages of Spanish-speaking Participants.	82
3.6	Years Spent in a Spanish-speaking Country.	82
3.7	Spanish Lexical Proficiency Results.	86
3.8	Spanish Lexical Proficiency: Time to complete task.	87
3.9	Self-reported proficiency measures: Spanish.	89
3.10	Correlation of Lexical Task with Self-Reported Measures: Spanish	91
3.11	Correlation of Lexical Task Time with Self-Reported Measures: Spanish	92
3.12	Spanish Eye-tracking Data.	95
3.13	Spanish Results: By-participant results.	98
3.14	Interaction between Condition and Trial: Controls.	100
3.15	Interaction between Condition and Trial: Heritage.	101
4.1	Polish Match Condition Display.	127
4.2	Polish Mismatch-M Condition Display.	127
4.3	Polish Mismatch-N Condition Display.	128
4.4	Polish Auditory Stimulus Frame.	129
4.5	Ages of Polish-speaking Participants.	132
4.6	Years Spent in a Polish-speaking Country.	132
4.7	Polish Lexical Proficiency Results.	135
4.8	Polish Lexical Proficiency: Time to complete task.	136
4.9	Self-reported proficiency measures: Polish.	137
4.10	Correlation of Lexical Task with Self-Reported Measures: Polish	139
4.11	Correlation of Lexical Task Time with Self-Reported Measures: Polish	140
4.12	Polish Eye-tracking Data: Feminine targets.	143
4.13	Polish Eye-tracking Data: Effect of trial number (feminine target).	145
4.14	Polish Eye-tracking Data: Masculine targets.	145
4.15	Polish Eye-tracking Data: Effect of trial number (masculine target).	147

4.16	Polish Eye-tracking Data: Neuter targets.	148
4.17	Polish Eye-tracking Data: Effect of trial number (neuter target).	150
4.18	Polish Proficiency: Effect of vocab time on condition.	152
4.19	Heatmaps illustrating the three-way interaction between proficiency (measured by vocab time), condition, and trial number. Bluer shades indicate faster looking times on the corresponding mismatch condition, while redder shades indicate slower looking times.	154
5.1	Proficiency comparison: Percent correct of vocabulary items.	168
5.2	Proficiency comparison: Time to complete vocabulary task.	168
C.1	Spanish: Random slopes for feminine target items.	215
C.2	Spanish feminine target item: Interaction between condition and trial number.	215
C.3	Spanish: Random slopes for masculine target items.	218
C.4	Spanish masculine target item: Interaction between condition and trial number.	218

DLA BABCI I DZIADKA.
DLA MAMY I TATY.
DLA ADASIA I WITKA.

Acknowledgments

Thank you first and foremost to my committee.

To Masha: Thanks for taking me in and guiding me through the last six years. For telling me when I did a good job, and when I really needed to work on my writing – both were motivational. Thank you for pushing me to work hard, and for supporting taking time off for family, friends, and life events (both good and bad). I clearly would not be the linguist I am today (with a dissertation and a job!!) without your endless support.

To Jonathan: Thank you for adopting me this year. You were dropped into this halfway through the project, and yet you dedicated so much time and patience to learning about me and my work. I am incredibly grateful that various circumstances brought you to Harvard at just the right moment.

To Kevin: Thanks for being the person I could come to with my stats questions, big or small, from conceptual matters to issues of debugging code. Your guidance kept me on track and led to analyses and visualizations that I am very excited about.

To Holger: Thank you for joining this project so readily, offering your expert support, and pushing me to dig deeper into the data. Your support from the conceptualization of the project (now two years ago) has been instrumental in arriving at its current form.

This project could not have been executed without the financial support of the Institute of Quantitative Social Sciences at Harvard and the logistical support of the Harvard Center for Brain Sciences. And the project would be nothing without the engagement of the local community – the participants in this study. Thank you for your time and enthusiasm, and for sharing your experiences as heritage speakers.

A pryoone who roject of this scale is really a group effort. I would like to thank anyone and everyone who has provided help or support along the way. Thank you to Sadie and Michelle, my two top-notch RAs. To Stephanie McMains, who is the MATLAB god at the Center for Science at Harvard. To Esther Rinke, Héctor Álvarez Mella, Marta Rudka, Vanesa Rodriguez, and the rest of the folks at the University of Heidelberg for supporting me while conducting the German portion of this study. Thank you to Irina Sekerina and Bernhard Brehmer, who offered their expertise. Thank you to Gigi Luk, Ran Wei, and Tiffany Yang, who were so generous with their eye-tracking equipment and software prior to my trip to Germany. Thank you to Ewa Brzezińska and Kacper Kubień for helping develop the Polish materials, and to Juan Arias for helping develop the Spanish materials. Thank you to Hania Bigoszewska for lending her artistic talents.

Writing this dissertation overlapped a great deal with going on the academic job market, and so

I want to thank those who helped me in the job application process. Without them, I would be much less sane, and the outcome of both the dissertation and the job market would have been less favorable. I'm especially grateful to Omer Preminger and Jenneke van der Wal, who wrote letters of support for my applications. Many others shared advice and their own written materials that helped guide my writing and other preparation. In no particular order, thanks to Francisco Moreno, Kate Davidson, Ruth Kramer, Terje Lohndal, Mark Norris, Greg Scontras, Lauren Clemens, and Rajiv Rao. Thank you also to Adam Beaver and Pamela Pollock from the Harvard Bok Center for Teaching and Learning for their support in preparing my teaching portfolio. And of course, thank you to the folks at Iowa – I cannot wait to join you all in the fall.

Thank you also to those who have supported me not just in this last year but more generally in my academic endeavors. Thank you first and foremost to John McWhorter. This whole linguistics thing is your fault, so thanks for sticking with me even after I left Columbia. Thank you also to Bert Vaux, whose supervision in my time at Cambridge a long long time ago was instrumental in my choice to go to graduate school. Thank you also to Dean James Valentini (Deantini) for support in all things academia, and to Dean David Madigan for supporting my enthusiasm in bringing more statistics to the linguistics world.

Thank you also to my fellow linguistics graduate students and postdocs, who have made the last few years so much for more. Thanks to my cohort: Dora and Lena (and Tyler). We did it! Thank you also to Michelle Yuan, Athulya Aravind, and Masoud Jasbi, as well as Cristina Aggazzotti, Paulina Lyskawa, and Carina Kauf for your friendship both in academic contexts and beyond.

A big thank you to my family, who have supported me even though I decided to study something that is much less marketable than statistics, and for being there every step of the way. Our family's story is the inspiration for my research interests. And thank you to the Polish Clique. Unbeknownst to you all (or me!) you were my first heritage speaker guinea pigs. If I may say, we're a pretty amazing group of heritage speakers.

A few personal thank yous are also in order. First and foremost to Robert: I cannot fit into this space how grateful I am for everything. Thank you for being there through the good and the bad. I know you'll say that it's not true that I couldn't have done it without you... You're probably right, but I would have been a lot worse for wear. Thanks to Michelle and Amy for our endless g-chats and for our annual trips. To my New York squad – Katey, Henrique, Nick, Susan, Jimmy, and co. – for being ready to hang out whenever I needed to escape the my favorite city. To the rest of the Ellery squad for making graduate school the opposite of lonely and boring. To Emily, Brent, and Jack for their friendship since high school. To Marcella at Cambridge Coaching, for her trust and support, and for the job that gave me much more financial stability than I was anticipating in grad school.

And thank you to heritage speakers everywhere. Every one of us has a slightly different story, united by complex relationships to our heritage language(s) and heritage culture(s). I hope that through my work I can shed light on even the tiniest slice of what it means to be a heritage speaker.

0

Introduction

The topic of study for the current project is grammatical gender in Heritage Spanish and Heritage Polish, bringing together theoretical understanding of gender as a syntactic feature and the rapidly growing field of heritage linguistics, additionally informed by work in psycholinguistics and language acquisition. There is a general consensus in the literature on agreement errors in heritage languages that heritage speakers consistently demonstrate difficulty in gender assignment and agreement. This motivates two broad questions for work on gender in heritage languages:

1. What do heritage speakers know about grammatical gender in their HL?
2. How do heritage speakers deploy their knowledge of grammatical gender in online tasks?

The first question is addressed for Spanish in work by [Scontras et al. \(2018\)](#), in which the authors compare heritage and control speakers of Spanish in an acceptability judgment task and determine that the internal structure of gender is the same in both grammars. However, they also propose that the relationship between number and gender is different in the heritage and monolingual grammar (see the discussion in Section 1.4).

The present dissertation contributes to the literature on gender in the heritage grammar by addressing the second question, how heritage speakers use the gender feature in online tasks. The emphasis is on gender in the nominal domain, as some work discussed above suggests gender agreement errors are more common in the nominal domain than in the verbal domain. Since there has been ample work done on gender in (Heritage) Spanish, the first study presented (Chapter 3) considers this question for heritage speakers of Spanish, fitting into an existing body of literature but extending it in a meaningful way by employing a new, more sensitive methodology. The eyetracking methodology employed in this study is an implicit measure of linguistic ability, giving more confidence that we can assess heritage speakers' true linguistic ability, unencumbered by the task effects involved in some studies that rely on explicit knowledge or metalinguistic awareness of the heritage language. What we find in Chapter 3 is that despite broadly recognized surface issues in production and comprehension of gender, heritage speakers of Spanish can use gender to facilitate lexical retrieval in online processing in a monolingual-like manner. Couching these results in the context of similar studies on monolingual, child, and L2 speakers of Spanish allows them to be interpreted as evidence in support of treating heritage language acquisition as closer to L1 acquisition, despite the fact that heritage language acquisition shares with L2 acquisition being nonconvergent and dependent on sociolinguistic factors.

Recognizing that, crosslinguistically, the gender feature is not restricted to the binary masculine-feminine distinction seen in Spanish, as well as many other Romance languages, in Chapter 4 I perform the same study on control and heritage speakers of Polish, which has three gender values: masculine, feminine, and neuter. This study again bears on the second broad question stated above, regarding the use of grammatical gender in online tasks, and results indicate that heritage speakers of Polish can also use gender to facilitate lexical retrieval in online processing in a monolingual-like manner. Moreover, the syntax of Slavic languages (including the gender feature) has not been studied nearly as extensively from the psycholinguistic perspective as the syntax of Spanish and other Romance languages. Thus the results of the study contribute novel experimental data in support of a hierarchical analysis of the internal structure of the gender feature in Polish, both in the monolingual grammar as well as in the heritage grammar – which in turn bears on Question 1 above.

A significant feature of these two studies is that they were conducted using the same materials and the same experimental design (to the extent that the syntax of the languages allowed). This allows for a direct comparison of the performance of heritage speakers across two languages. As noted previously, such systematic comparisons are largely absent from the field of heritage linguistics. However, as Chapter 5 discusses, such comparisons and the microvariations they reveal suggest asymmetries between various syntactic elements that draw implications and motivation for further study. The comparison in Chapter 5 suggests possible interpretations of these asymmetries and proposes follow-up studies that might adjudicate between several possibilities, leaving open exciting questions regarding the nature of the use of gender information on determiners and modifiers.

Additionally, the studies presented here were designed with a further research question in mind: What is the role of transfer in the development of the heritage grammar (a question introduced in Section 2.5). This is a very relevant question for the study of gender: the heritage populations tested here have English as their dominant language. By most accounts, English does not have grammatical gender, and so it is largely assumed that heritage speakers' error rates with gender may be at least

to some degree attributed to the lack of a gender feature in their dominant language. The implication of this line of thinking is that a heritage speaker whose dominant language has gender may not present similar difficulties in gender assignment and agreement. What is missing from the field is a systematic comparison of heritage speaker populations with different dominant languages that might confirm or reject this hypothesis for gender, and so the experimental materials here were designed in order to subsequently test Heritage Spanish and Heritage Polish populations in Germany (German has three genders: masculine, feminine, and neuter). The stimuli are balanced in such a way that the gender of the German equivalents of the stimuli is also balanced throughout the study, in a way that will allow the results of running the studies in Germany to be interpreted in the context of transfer. At this time, the study has been conducted in Germany although the results have not been analyzed. Ultimately this systematic comparison across four heritage speaker populations will be – to my knowledge – the first of its kind in the field of heritage linguistics. Until then, the comparison of Heritage Spanish and Heritage Polish here is intended to set a precedent for more frequent comparisons not only within but across heritage speaker populations.

I provide here a brief outline of the following pages:

Chapter 1 introduces the notion of grammatical gender as it is defined and studied in the linguistic and specifically syntactic literature. This chapter provides an overview of the crosslinguistic properties of the gender feature and provides some discussion of the notion of “default” gender, which will be relevant to later chapters. The chapter also outlines some additional questions of interest in the body of research on gender, which the empirical results in the dissertation will not bear on these questions, although certain planned extensions of the studies may do so.

Chapter 2 provides a broad overview of the field of heritage linguistics, touching on work from various subfields of linguistics to illustrate the rapidly expanding interest in heritage speakers as a window onto acquisition and as a testing ground for linguistic theory. This chapter also introduces topics of interest in heritage linguistics that motivate choices in the experimental design and popula-

tions made in the empirical chapters.

Chapter 3 presents the eye-tracking study conducted on heritage and control speakers of Spanish. This chapter discusses various design choices and the development of the experimental materials and procedure, all of which carry over to the Polish study presented in the subsequent chapter. The results of the study indicate that heritage speakers of Spanish can use gender on articles to facilitate lexical retrieval, although they are slower to do so than the control group.

Chapter 4 extends the study in Chapter 3 to a language with a three-valued gender feature – Polish. Heritage and control speakers were tested using the same experimental design and procedure as in the Spanish study, with the major difference being that prenominal adjectives – not articles – are the locus of the gender feature used to make predictions about the subsequent noun. Here again, results indicate that heritage speakers can use gender to facilitate lexical retrieval. Asymmetries in how the gender of the distractor impacts participants' ability to use a given gender in the task suggest a hierarchical analysis of gender for both heritage and monolingual Polish.

Chapter 5 considers in more detail a contrast between the Spanish and Polish results. While heritage speakers of Polish perform the task just as quickly as the control group, heritage speakers of Spanish are slower than the control group. This chapter considers several options for why this might be the case, ultimately emphasizing the difference between accessing gender information on determiners versus on modifiers. The chapter ultimately leaves open the question of what key property lies at the core of this difference and suggests several follow-up studies to test some of the candidate hypotheses.

1

Grammatical gender

THE GOAL OF THIS CHAPTER is to provide a broad overview of grammatical gender as an object of linguistic study, specifically in the syntax of the nominal domain. For languages with grammatical gender, every noun must be assigned to a gender category based on some set of assignment rules. This gender feature then affects the behavior of associated words – other nominal elements that inflect to match the gender feature assigned to the noun. In Section 1.1 I introduce this notion, and

then consider the various ways in which it might be instantiated crosslinguistically in Section 1.2. In Section 1.3 I discuss the notion of default gender and identify which of the many versions of “default” are of interest in this dissertation, and in Section 1.4 I provide background on other questions of interest regarding gender in the syntactic literature that are nevertheless outside the scope of this dissertation.

1.1 WHAT IS (GRAMMATICAL) GENDER?

Before considering the possible analyses of gender, it is first important to define the notion of gender as it is used in this dissertation. In languages with grammatical gender, all nouns are assigned a gender based on a system that is independent of any real-world properties and therefore arbitrary, as illustrated in (1) (this is in contrast with *notional gender*, which can be determined based on real-world properties, such as sex or animacy, of a noun’s referent). Thus, we can define gender as a method of grouping nouns according to their morphological shape and/or syntactic behavior. Defined this way, the gender category of a noun can reliably be determined only by the shape of associated words (Hockett 1958:231), not necessarily by the morphophonological form of a given noun, although in many languages there are morphophonological correlates or cues to gender category on the noun itself.

- (1) Spanish
- a. *la manzana* ‘the apple, fem.’
 - b. *el libro* ‘the book, masc.’

The “behavior of associated words” that allows us (and the learner) to determine the gender category of a noun includes distributional evidence, such as the form of the determiner that co-occurs with nouns in a particular gender category. In German, which has little to no gender correlates on

nouns, (singular) masculine nouns occur with the determiner *der*, feminine nouns occur with the determiner *die*, and neuter nouns occur with the determiner *das*.

(2) German

- a. *der Apfel* ‘the apple, masc.’
- b. *die Uhr* ‘the clock, fem.’
- c. *das Buch* ‘the book, neut.’

Similarly, a reliable indication of the gender category of a noun is the syntactic behavior of other associated words, specifically the inflectional morphology on nominal modifiers (this phenomenon is commonly referred to as concord) and verbs. In (3) the shape of the modifier varies with the gender of the noun (this example also repeats from above the distributional facts regarding articles *el* and *la* in Spanish). The class of associated words that can reveal the gender of a noun in some languages also includes verbs, as in the Polish examples in (4).

(3) Spanish

- a. *el papel blanco* ‘the white paper, masc.’
- b. *la sal blanca* ‘the white salt, fem.’

(4) Polish

- a. Stół stal na dywanie.
table.M stand.PST.3SG.M on rug
‘(The) table stood on (the) rug.’
- b. Sól stal-a na blacie.
salt.F stand-PST.3SG.F on counter
‘(The) salt stood on (the) counter.’

At this point it is important to distinguish between two notions: gender assignment and gender agreement. The former is the system by which a speaker decides what gender category a noun be-

longs to. The latter is the process by which that gender category information is spread to associated words in the syntax. This distinction will be relevant in subsequent sections and chapters. While gender assignment rules will be mentioned for both Spanish (Chapter 3) and Polish (Chapter 4), the focus of this dissertation is not on how speakers determine the gender of a noun but how they use gender agreement (concord) on other elements in the nominal domain to anticipate the noun itself.

Gender assignment is sometimes conflated with declension class – a concept that is relevant particularly for Polish, but a related concept of *word marker* also surfaces in Spanish (Harris (1991)). Polish, like other Slavic languages, has a nominal case system that determines the inflectional ending of a noun based on the role it plays in the clause. Nouns in a single declension class pattern together in that they take the same set of inflectional endings. It just so happens that declension class and gender are heavily correlated in Polish (as in many other languages): nouns whose nominative inflectional ending is *-a* are typically feminine, those whose ending is *-o* or *-e* are typically neuter, and those that end in a consonant are typically masculine. However, declension class is only one of the factors that goes into computing gender assignment – notional gender and other aspects of the phonological form of the word may also play a role. Thus, not all nouns in a declension class will be of the same gender. For instance, nouns may be in the typically feminine declension but show masculine agreement (as with *artysta* ‘artist’). More exceptions will be discussed in Chapter 4. Because this dissertation will only focus on nouns with “canonical” endings (those that belong to the same declension class as most nouns in a particular gender category), I will say that *-a* is the “typical feminine ending” in Polish, but with the understanding that this is shorthand for a correlation between gender and just one of the notions that is involved in mapping nouns to gender categories.

For the remainder of this dissertation, unless stated otherwise, gender refers to the abstract feature that groups nouns according to their morphological shape and/or syntactic behavior, as discussed above. Examples in the discussion so far have been drawn from languages with two genders (masculine and feminine, as in Spanish) and three genders (masculine, feminine, and neuter, as in

German and Polish), but to illustrate the power of gender as a categorizing feature we can consider other gender divisions in the following section.

1.2 GENDER CROSSLINGUISTICALLY

Crosslinguistically, the gender feature can vary along several axes. Here, I will discuss variation in terms of complexity (the number of gender feature values) and in terms of what types of properties determine membership of nouns in the language to certain gender categories. This dissertation focuses on gender in Spanish and Polish – both European languages – it is important to bear in mind that the systems investigated in the project are only two of many, and that the gender divisions and their labels are not meant to be representative of gender in all its crosslinguistic realizations.

The first axis along which languages exhibit variation with respect to the gender feature is the number of gender values. Of course some languages, such as English, do not have the notion of grammatical gender (although even English has gender in the pronominal system). Most Romance languages have two genders, while Latin and some other modern European languages have three. Here the common gender distinctions are masculine versus feminine (as in Spanish, French, and Italian) and masculine versus feminine versus neuter (as in Polish, German, and Russian), but even within Europe the labels can vary from ones traditionally associated with “gender”, as Swedish and Danish have two-gender systems where the gender categories are labeled “common” and “neuter”. However, it is important to note that, for the most part, “nothing rests on the actual labels used” (Corbett 1991:9). Abstracting away from natural gender and thinking of gender as a syntactic feature in the manner described above, the labels could just as easily be “Category A”, “Category B”, etc. (although in some cases this would lose some generalizations that might be made about a semantic core of these systems, which will be discussed below).

As an abstract categorizing feature, gender is not restricted to having two or three values. Lan-

guages can have far more genders, with the Bantu language family being particularly well known for having a complex gender system. Bantu languages have on average 12-20 noun classes that combine number and gender information. In (5) below, the pronominal prefix, the suffix on the demonstrative, and the prefixal subject marker on the verb are all determined by the noun class of the head noun. More precisely, *-toto* ‘child’ belongs to noun class 1 (decomposed into Gender A, singular following the system in Carstens (1991)), and so in (5-a) the head noun has the prefix *m-*, the demonstrative enters into concord and takes the suffix *-yu*, and the verb takes the prefix *a-*. However, the plural ‘children’ belongs to noun class 2 (Gender A, plural), and the analogous elements are different (5-b). Similar effects can be seen for *-tabu* ‘book’ which in the singular is class 7 (Gender D, singular) (5-c) and in the plural is class 8 (Gender D, plural) (5-d).

(5) Swahili

- a. m-toto hu-yu a-na-soma
1-child DEM-1 ISM-PRES-read
‘This child reads.’
- b. wa-toto ha-wa wa-na-soma
2-children DEM-2 2SM-PRES-read
‘These children read.’
- c. ki-tabu hi-ki ki-na-som-wa
7-book DEM-7 7SM-PRES-read-PASS
‘this book is reading.’
- d. vi-tabu hi-vi vi-na-som-wa
8-book DEM-8 8SM-PRES-read-PASS
‘These books are reading.’

For those whose experience with gender has been only in the context of Romance languages, referring to a system with 10+ noun categories as “gender” may be difficult to reconcile with the Romance/European system of masculine and feminine (and neuter) that seems to be rooted in some real-world distinctions. However, returning again to the notion that gender is a way of grouping

nouns according to syntactic behavior, it should be unsurprising that there is more than one way of dividing nouns into categories.

This brings us to the other axis along which languages show considerable variation: what properties determine the division of nouns into gender categories. [Corbett \(1991\)](#) provides a broad overview of these divisions, demonstrating that languages use semantics, morpho(pho)nology, or both to categorize nouns. In semantic systems, semantics is the only factor that determines gender category membership in the language. Sex-differentiability is one type of factor – for instance, Kala Lagaw Ya differentiates nouns referring to male-denoting nouns from all other nouns, while Diyari differentiates female-denoting nouns from all others. [de La Grasserie \(1898\)](#) provides a list of other semantic factors along which languages may make divisions (into two categories, or into more, using a combination of these factors): animate/inanimate, rational/non-rational, human/non-human, strong/weak, augmentative/diminutive, male/other, male-human/other, male/female/non-sexed. [Corbett \(1991\)](#) argues that this list is incomplete, calling up languages such as Andi, which has a gender for insects, and Dyirbal, which has a gender for non-flesh food.

In formal gender systems, divisions into noun categories appear to be divorced from semantic notions such as the ones above – in other words, it would be difficult to identify semantic rules that would be reliable predictors of gender category membership. Instead, morpho(pho)nological properties determine gender. However, as [Corbett \(1991\)](#) notes, there is always a semantic core of nouns whose category can reliably be predicted from their meaning or properties of the real-world referent. The remaining “residue”, as Corbett refers to it, is distributed into gender categories based on formal morphological and/or phonological properties. Most of the languages discussed to this point fall into this category. It is certainly true for Romance and Slavic languages, in which sex-differentiable nouns are typically reliably categorized as masculine or feminine, while the remainder of nouns (most of them inanimate) are categorized according to the morphophonological shape of their ending. When semantic and morphological cues compete, formal cues usually win (ex. Polish

dziewczę ‘maiden’ is semantically feminine but has an ending typical of neuter nouns and determines neuter agreement). Languages vary with respect to whether semantics or morphology takes precedence over the other.

With so many different gender systems to choose from, why study Spanish and Polish? The empirical domain of the studies presented here is gender in heritage languages. In studying heritage languages, it is important to choose a language in which there is existing work on the phenomenon of choice in the monolingual grammar – this isolates the variable of interest from issues of language contact and bilingualism, before delving into these matters in examining the corresponding heritage language. Thus, Spanish is the ideal starting point because its gender system has been relatively well-studied in the formal literature, the psycholinguistic literature, the acquisition literature, and even to some extent in the heritage literature, at least with respect to production and comprehension errors. Additionally, Heritage Spanish populations are easily accessible in many regions of the US, which presents a very practical motivation to study this heritage language. Of course, Spanish is an entry point: choosing a two-gender system allows us to investigate the gender in the chosen empirical domain in its simplest (binary) instantiation is a first step that allows us to set some expectations for studies of more complex systems. Specifically, Polish allows us to add one layer of complexity to the investigation: Polish is a three-gender system, but like Spanish it is a mostly formal/morphophonological, with a semantic core determined by sex-differentiability. Unlike Spanish, Polish gender is not nearly as well studied, but enough formal and acquisition work exists to have confidence in interpreting the experimental results, and making a substantial contribution to a body of literature that is generally in need of input from psycholinguistic studies.

1.3 INTERNAL ORGANIZATION OF GENDER FEATURES AND THE NOTION OF “DEFAULT” GENDER

Research on gender is also concerned with the relationship of gender values to each other within a single language, and particularly with the notion of “default” gender (Corbett & Fraser (1999), Haspelmath (2006)). A default gender occurs when no other rules apply. It is inherently a relative notion, as what is considered default is dependent on whether we consider gender assignment or gender agreement, and on which language we examine. Crosslinguistically there are a variety of conditions in which the need for a default may arise – which conditions invoke the default varies from language to language. We identify the default gender based on which gender category arises in these conditions, and it is important to note that there is *a priori* no way to determine which gender this may be. For instance, in a two-gender system (as in Spanish, which will be discussed further in Chapter 3), either gender is a viable candidate for the default; ultimately the default can only be determined by considering conditions in which default gender assignment or default gender agreement arise.

With respect to gender assignment, in determining the gender category of a noun in their language speakers may use some combination of notional and morphophonological cues (such as declension class, discussed above). The default arises when none of the usual gender assignment rules apply, what Corbett & Fraser (1999) refer to as “too little information”. In a semantic gender system, this may occur when the gender of the referent is unknown. In many instances, neologisms or loanwords are considered good test cases for the default, as their morphophonological form is usually distinct from that of most noun in the language. Although, work on English loanwords in Polish by Fuchs (2014) discussed in Section (4) reveals that loanwords are not always a reliable indicator of the default gender.

With respect to gender agreement there are several other common conditions in which a default

may need to be invoked. One such condition may arise when the gender of the controller of agreement is not specified, as with non-prototypical noun phrases such as infinitives (what [Corbett & Fraser \(1999\)](#) refer to as “neutral agreement”). In a similar vein, the gender of the controller may be underspecified, as when referring to a child, also requiring default gender agreement. In some languages, a default gender may also be invoked when there is more than one gender cue competing to determine gender agreement, as in the case of conjoined noun phrases, although some languages do have rules for computing gender in this conditions, such as closest conjunct agreement.

The notion of default has been used to refer to various concepts in the literature (see [Haspelmath \(2006\)](#) for full discussion), and as [Corbett & Fraser \(1999\)](#) and [Haspelmath \(2006\)](#) note, even within a single language the choice of default strategies across conditions need not correlate. It is therefore important to establish which notion of default is being assumed. In this dissertation, the empirical domain concerns how gender information on articles and adjectives is accessed in online tasks, and so the primary interest is in gender agreement, and where default agreement may arise.

One final note regarding defaultness has to do with [Haspelmath \(2006\)](#)'s notion of a textual default, or frequency. This is specifically frequency of use in oral and written speech, and not necessarily frequency of occurrence in the real world. This notion is relevant because, as Haspelmath argues, it may be the root cause for other types of markedness/defaultness that occur in the literature, such as what Haspelmath refers to as “cognitive markedness”, according to which a non-default value or construction may entail a larger cognitive load and therefore take more time to process. Frequency need not correlate with all types of defaults, but in light of this, the frequency of relevant constructions and words will be addressed and controlled for where possible.

1.4 THE RELATIONSHIP BETWEEN GENDER AND NUMBER

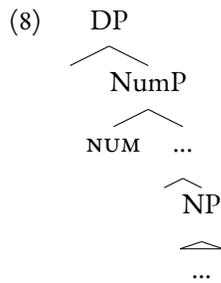
In addition to being concerned with the relationship of the different values/categories within the gender feature, research on gender is also interested in how the gender feature interacts with other features in the nominal domain, particularly with number. I will provide a brief overview of the issue here, although the studies presented here will not bear on the nature of the relationship between number and gender, with the exception of some discussion of a possible extension study in Section 3.9.

Crosslinguistically, languages regularly combine number and gender information on a single morpheme. In Italian, number and gender are fused in a single suffix (6). In German, for a particular subset of nouns, plurality is indicated only through change in the definite article (which is known to carry gender information), with no change to the noun itself (7). Bantu noun classes, discussed above in Section 1.2 also expone number and gender simultaneously.

- (6) a. *libro* ‘book, masc.’
libri ‘books, masc.’
b. *mela* ‘apple, fem.’
mele ‘apples, fem.’

- (7) *das Mädchen* ‘the girl’
die Mädchen ‘the girls’

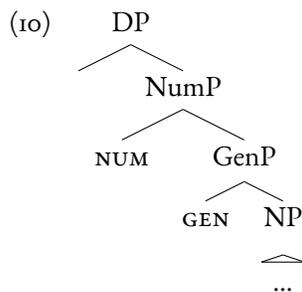
These facts have motivated research regarding the structural relationship of the two features. There is ample evidence for the number feature, which has consistently interpretable semantics, to head its own projection, and it is now commonly assumed that NumP is universally projected below DP in the syntax of the nominal domain, as illustrated in (8) (for detailed discussion, see Ritter (1991, 1993), Longobardi (1994, 1996)).



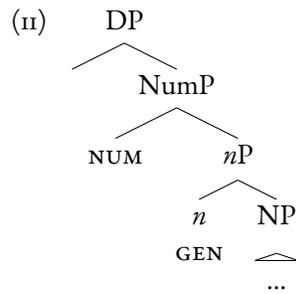
If we assume that gender must also be represented in the syntax, given that it participates in syntactic processes such as agreement, the question then is how gender is structurally related to number. One option is that it might be bundled with number information on NumP. Ritter (1993) argues in favor of this analysis with particular emphasis on Romanian, in which nouns sometimes take on a different gender in the plural:

- (9) *un glas* ‘a voice, masc.’
două glas-uri ‘two voices, fem.’

The competing hypothesis* is that gender is represented independently, split from number. Two possible instantiations of this are an analysis in which gender heads its own projection GenP (10) (Picallo (1991), Bernstein (1993)), and an analysis couched within the framework of Distributed Morphology in which gender is located on *n* (n) (Kramer (2014, 2015)).



* An additional competing hypothesis is that gender is located on the root, although the body of literature on the syntax of the nominal phrase has largely moved away from this analysis.



Split analyses of number and gender commonly argue that number and gender are often represented on independent morphemes, and that when they are, they have a consistent morpheme order: STEM-GEN-NUM. [Kramer \(2015\)](#) shows that a split analysis of gender better accounts for these languages, and that even languages such as Italian, Walloon, and Romanian that combine gender into a single morpheme can be captured under a split analysis of gender. In fact, Kramer argues that the split analysis of number and gender is universal.

While most investigations on this topic have been formal, [Fuchs et al. \(2015\)](#) conducted an experimental study on monolingual Spanish, in which they presented participants with auditory stimuli with manipulated errors in agreement in number, gender, or both features between the noun and the predicate in agreement attraction constructions. They found that errors in both features appeared to have an additive effect, being rated as less grammatical than sentences with a single agreement error. They argued that this constituted evidence that number and gender were independent agreement processes, and should therefore be represented independently in the syntax.

However, in a follow-up study on Heritage Spanish, [Scontras et al. \(2018\)](#) challenge Kramer's claim that the split analysis of number and gender is universal. Heritage speakers perceived the researchers' stimuli with two agreement errors as equally degraded as the stimuli with only one agreement error. The authors took this to be evidence that valuation in number and gender is a single process in the heritage grammar, suggesting that Heritage Spanish bundles number and gender on a single projection.

While this dissertation will not directly address the issue of how gender is structurally related to number, it will suggest an extension study that could address this issue in Section 3.9. As will be discussed, [Scontras et al. \(2018\)](#) make certain predictions regarding the opacity of a bundle of features that could be tested in the experimental paradigm presented in Chapter 3. Such an investigation would have implications for whether the bundled hypothesis is appropriate for the grammar of Heritage Spanish and thus whether a universal split analysis of number and gender can be maintained.

1.5 WRAP-UP

Grammatical gender is a very rich topic – much has been written on the issue of gender, and I certainly cannot do justice here to this rich field. The above discussion is intended to provide a general overview of the concepts that are relevant to the studies and implications made in the subsequent chapters. Chief among these concepts is the notion of gender as a categorizing feature, and how this feature is assigned to nouns and then spread to associated words in the nominal phrase. While going forward I will refer to noun endings as indicative of a particular gender feature, this chapter discussed the fact that morphophonological form (including declension class, relevant in Polish) is only one of several factors that might be taken into account in gender assignment.

In the studies described in Chapter 3 and Chapter 4, gender on prenominal elements (specifically articles and adjectives) will be investigated in light of how speakers use this information in online tasks to anticipate subsequent nouns. This falls under the scope of gender agreement, and so default gender agreement is also addressed in the relevant places. More broadly, and awareness that grammatical gender crosslinguistically varies from binary distinctions (such as masculine/feminine in Spanish) through systems of 20+ noun classes as in Bantu languages. This awareness motivates using Spanish as a starting point, but also motivates moving to languages with more complex gender system. The first step in this move is to investigate the three-gender system in Polish, as I do here,

but studies of other gender systems would be informative in developing a broad, generalizable understanding of how (heritage) speakers use gender to facilitate lexical retrieval.

2

Heritage languages & linguistics

HERITAGE LINGUISTICS IS A RAPIDLY GROWING FIELD. The first reason for this is a very practical one: demographic shifts have led to an increase in the number of bilinguals who qualify as “heritage speakers”, and who enter language learning classrooms at the high school and college level hoping to learn to speak their home language more like their parents. A need to develop appropriate curricula and teaching materials drives a need for scientific investigation of what the linguistic knowledge and

ability of these speakers really are. This demographic shift has also led to a growing interest from the side of linguistics, which as a field has begun to recognize the importance of developing theories that account for phenomena as they are instantiated not only in monolingual grammars but also in grammars under pressure from language contact. In addition, given that heritage languages develop under conditions of reduced input to the heritage language (as will be discussed below), heritage linguistics gives a window onto questions of language acquisition, in particular those questions concerning the role of input in the development of the grammar. The purpose of this chapter is to provide a general overview of the target population in heritage linguistics, and how and why these speakers are of interest to the broader linguistic community.

Data from the US Census Bureau shows that in 2014/2015, there were 42.4 million immigrants in the United States, making up a record 14% of the country's population. The Pew Research Center reports that a further 12% of the American population is 2nd generation immigrants – those who were born in the United States and have at least one foreign-born parent. In American public schools, 23% of students (10.9 million students) come from immigrant households. These students and their families often have perfect or imperfect command of the language spoken in their home country, in addition to English. As will become clear below, these are prime candidates for being considered heritage speakers.

Multilinguals make up most of the population of the world. Even in a country like the United States – often perceived as a monolithic monolingual country – over a quarter of the population speaks more than just one language. In recent years, linguistic theory has begun to reflect this truth about the status of mono- vs multilingualism in the world, recognizing that the study of *human* language ought to be more nuanced and complex in order to account for the very human ability to maintain multiple grammars.

At the outset of studies in generative grammar, Chomsky (1965) set the standard as such: “Linguistic theory is concerned primarily with an ideal speaker-listener, in a completely homogeneous

speech-community, who knows its language perfectly...” (Chomsky 1965:3). Much of the theory developed in the decades since then has relied on judgments and performance evaluations of speakers who fit this rather idealized mold: monolinguals. But as has been observed by many researchers, linguistic theory has made enough progress that it can now move beyond monolinguals and tackle the unique set of challenges posed by speakers who claim competence in more than one language (Montrul (2010a), Benmamoun et al. (2013b), Lohndal (2013), Polinsky (2018), a.o.).

As a subgroup of multilinguals, heritage speakers have sparked a particularly exciting body of research. Recall the 12% of the American population that is classified as second-generation immigrants: born in the United States with at least one foreign-born/immigrant parent. These individuals often have some ability in the language of their foreign-born parent(s). Proficiency in the parents’ language varies considerably from case to case depending on a range of factors, and more and more young adults with such a linguistic background look for resources in their educational system that will help them regain command of language they learned at home and maybe subsequently lost to some degree. These heritage speakers are a decidedly heterogeneous group, but their unique linguistic abilities have generated a substantial line of research with implications for the field’s understanding of theory and of acquisition. The following section will discuss in more detail the profile of a heritage speaker before returning to the motivation for studying the linguistic abilities of heritage speakers and the impact these studies have had on the field.

2.1 WHO ARE HERITAGE SPEAKERS?

One of the first and most influential definitions of a heritage speaker is given by Valdés (2000): heritage speakers are “individuals raised in homes where a language other than English is spoken and who are to some degree bilingual in English and the heritage language.” This English-centric definition is a reflection of the fact that much of heritage language research – at least at the outset –

has been conducted in the United States, where English is indeed the dominant language – the one spoken by the heritage speaker in school or, more generally, outside of the home.

The field has since abstracted away from emphasizing English as the dominant language, but two key aspects of Valdes' original definition persist as the defining characteristics of a heritage speaker. First, Valdes considers to be heritage speakers those who were “raised in homes where a language other than English is spoken.” Heritage speakers are those who acquired their heritage language (abbreviated henceforth as HL) at home, before or simultaneously with English. While this formulation assumes a heritage language is spoken in the United States, more updated versions of this definition broaden the notion of a heritage language and heritage speaker to account for the fact that heritage languages arise not only in English-speaking communities.

A language qualifies as a heritage language if it is a language spoken at home or otherwise readily available to young children, and crucially this language is not a dominant language of the larger (national) society... [A]n individual qualifies as a heritage speaker if and only if he or she has some command of the heritage language acquired naturalistically..., although it is equally expected that such competence will differ from that of native monolinguals of comparable age. (Rothman 2009:156)

This definition encompasses heritage languages that develop outside of the United States. Kupisch & Rothman (2018) note that there are some differences between how studies of heritage languages have been presented in the literature that may imply outcome differences between heritage speakers in Europe as compared with those in the United States, but that this is confounded by various methodological factors. They argue that any such results should not be taken as an indication that “conditions for bilingualism vary across countries and continents” (Kupisch & Rothman 2018:575). Another more recent definition a heritage speaker echoes this emphasis on the acquisition of the heritage language as an LI, abstracting away from the geographical locus and identity of the majority language. Benmamoun et al. (2013b) state that a heritage speaker is...

...an early bilingual who grew up hearing and speaking the heritage language and the

majority language either simultaneously or sequentially in early childhood (roughly up to age 5; see [Schwartz \(2004\)](#), [Unsworth \(2005\)](#)) but whose L2 became their primary language at some point during childhood, typically after the onset of schooling. ([Benmamoun et al. 2013b:135](#))

Heritage speakers are thus a special case of childhood bilinguals. To a large degree, they acquire their HL much in the same manner that a child living in a monolingual community in the home country would – naturalistically and aurally – as reflected in the definition provided by [Rothman \(2009\)](#) above. This is clear-cut for those heritage speakers who grow up in the country of origin and only subsequently immigrate to the new country with their parents; these are clear cases of sequential bilingualism. For simultaneous bilinguals (ex. those with only one parent who speaks the HL) this is less the case. There are also in-between cases: children who grow up in monolingual households, but who are exposed to the majority language through television, family acquaintances, etc. Regardless, when heritage speakers enter the school system in their new country, they begin spending significantly more time in an environment that speaks the dominant language. The amount of input in the HL decreases drastically, and so heritage speakers begin to diverge from their monolingual counterparts in acquisition of the language. Any already acquired ability in the HL might even become weaker, depending on the context and degree of use of the HL, which may be impacted by the sociopolitical status of the language, whether it is tied to a religious community, and whether it has a written script ([Grosjean \(2008\)](#)).

The second key property that Valdes highlights is that heritage speakers are “to some degree” bilingual in the dominant language and heritage language. In this way Valdes acknowledges the spectrum of abilities that heritage speakers represent. A speaker’s proficiency in the HL can range from very low (a “receptive” speaker, who can understand the HL when it is spoken to her but who is reluctant to produce the HL because of severe limitations) to very advanced (performing like a monolingual in many or most domains of the language). Self-reports from heritage speakers confirm that, overall, heritage speakers – even when quite advanced – are more comfortable listening to than

speaking the HL (Carreira & Kagan (2011)).

So, who fits the description of heritage speakers? As suggested above, in most cases, heritage speakers are 2nd generation immigrants, but the definition can be broadened to include any children of families who “speak an ethnolinguistically minority language” (Montrul 2010a:3). Taken in this way, the concept of a heritage speaker can also include speakers of indigenous languages, as these speakers also become bilingual in the dominant language at an early age when they enter the school system and use the indigenous language in a limited number of contexts (Fishman (2006)). Their proficiency in the HL may be different than that of a purely monolingual speaker of the language. This should also be taken into account for those undertaking fieldwork whose informant might have been an early bilingual in the contact language (Scontras et al. (2018)).

Speakers in any of the groups listed above might claim imperfect competence in more than one language, but it is important that this discussion is not taken to suggest that they are not native speakers of their home language. There are two reasons to avoid calling heritage speakers “nonnative”. The first comes from language science: heritage speakers did acquire the heritage language as children, through first language acquisition – this is also how monolingual “native” speakers acquired the language. By the very nature of this type of acquisition that they experienced, heritage speakers are therefore a subtype of native speakers. Additionally, the term “native” speaker is tied to notions of cultural identity that are important to keep in mind, remembering that the term comes with cultural and sociological implications when used outside of strictly linguistic circles. Someone who identifies as a “native” speaker of a language is not only making some claim about their linguistic abilities but also identifying themselves as a member of some linguistic and cultural community. In recognition of this, the field of heritage linguistics has deliberately moved away from using “native” as a label that is in opposition to “heritage” in distinguishing between various speaker groups. A heritage speaker is a type of native speaker, and can be contrasted in linguistic performance with “monolingual”, “baseline”, or “control” speakers.

2.1.1 QUANTITATIVE MEASURES

Identifying who qualifies as a heritage speaker is crucial for identifying subjects for study in linguistic investigations of heritage grammars, and most researchers rely on a mix of demographic information and quantitative measures of proficiency to select their participants. The section above provided some demographic variables that delineate heritage speakers from other proficiency groups. The chief diagnostics are manner and length of exposure to the HL before onset of acquisition of the dominant language. Exposure to the majority language leads to lower proficiency in the HS during adolescence, while on the other hand children who have only exposure to the HL in a monolingual environment until age 8-12 are at a significant advantage for more advanced proficiency in the HL later in life (Montrul (2010a)).

Some experimental studies of heritage languages need to make more fine-grained distinctions within the population of heritage speakers, targeting or at least distinguishing between heritage speakers at different proficiency levels. Here, quantitative measures can be employed. For an experimental study, choosing a method of assessing proficiency can be a difficult task. Formal language tests are often used, but these often put heritage speakers at a disadvantage (as will be discussed below), and are time-consuming. Given this, various quantitative proxy measures of proficiency have been studied.

There are also several more quantitative methods for identifying heritage speakers and assessing their proficiency. Heritage speakers tend to have limited literacy in the HL, particularly if the HL and dominant language have different scripts, making written evaluations of proficiency unreliable. Speech rate has also been identified as a reliable indicator of heritage speaker ability in a language (Polinsky (2008a, 2011)). Speech rate is determined by observing spontaneous production and finding the word-per-minute output of the speaker. Polinsky (2008a, 2011) found that the speech rate of heritage speakers can be as low as 30% of the speech rate of monolinguals in the same language. The

significant slowdown has a principled source. “Lower proficiency speakers have difficulty accessing lexical items, which slows down their speech, but also significantly compromise comprehension and use of structure types that require quick incremental integration of information” (Benmamoun et al. 2013b:141).

As an additional diagnostic, lexical proficiency has been found to correlate with speakers’ control of the grammar (Polinsky (1997, 2000, 2006), O’Grady et al. (2009)). Asked to identify basic lexical items from a list of 200, those heritage speakers who were familiar with more words were later determined to make fewer errors in agreement, case-marking, etc. These results have been consistent across various heritage languages, including Arabic, Russian, Polish, Armenian, Korean, and Lithuanian (see Godson (2003), Albirini et al. (2013), a.o.), and they reflect a connection between lexical proficiency and grammatical knowledge; a similar connection has also been proposed for early child language acquisition (Fenson et al. (1994)).

2.1.2 DIFFERENT KINDS OF ACQUISITION

Heritage language acquisition is commonly thought of as being “between” first-language (L1) and second-language (L2) acquisition. This will be relevant in Chapter 3, where a comparison of the heritage Spanish results with previous work on child and L2 learners of Spanish will bear on whether heritage speakers’ ability to deploy knowledge of grammatical gender in online tasks places them closer to monolingual adults and children or to L2 learners. Here, I provide a brief overview of the important similarities and differences between the types of acquisition.

The fact that a similar relationship between lexical proficiency and grammatical knowledge has been proposed for both heritage speakers and child language acquisition reflects a common theme in the field of heritage linguistics that straddles two other subfields of linguistics: child first-language (L1) acquisition and adult second-language (L2) acquisition.

In some ways, the acquisition of a heritage language is much like the acquisition of a first lan-

guage by a child. Basic grammatical structures are acquired and mastered by monolingual children by age 3-4, as are relatively adult-like phonology and vocabulary (Guasti (2002), O'Grady (1997)) – at this age, the input to the heritage grammar is also primarily if not completely monolingual (this gets murkier in the case of simultaneous bilinguals). Importantly, this type of acquisition for both monolingual and heritage children is early and naturalistic, resulting from immersion in an environment in which the language is presented aurally. By contrast, L2 learners acquire the language predominantly from written materials, with access to metalinguistic information such as white space indicating breaks between words, and with far more exposure and practice with written registers of the language.

However, many parts of the grammar are not acquired until after age 4 or 5, including much higher-level vocabulary, complex syntactic structures, and many properties of pragmatics and the discourse-syntax interface. These are typically learned by monolingual children at school and through exposure to written registers (Chomsky (1969), Menyuk & Brisk (2005), Nippold (1998)). Since heritage speakers do not attend school in the heritage language, they typically do not have the chance to acquire or master these atypical, complex, or infrequent grammatical structures to the same degree that monolingual children do. For instance, Pascual y Cabo (2013) shows that heritage speakers of Spanish allow passivization of psych verbs, whose argument structure is mastered much later than that of standard agentive verbs. Psych verbs cannot be passivized in the monolingual grammar, suggesting that heritage speakers have not fully acquired the argument structure of psych verbs.

In this way, heritage language acquisition also resembles adult L2 language acquisition: its outcome is variable in proficiency and typically nonconvergent (Montrul (2010a, 2016)). Whereas monolingual language acquisition is convergent – in that monolingual children all eventually attain adult-like proficiency in the language – heritage language acquisition and adult L2 acquisition result in a spectrum of linguistic capability, ranging from receptive bilingualism to highly proficient speakers with monolingual-adult-like abilities. Also common to heritage acquisition and L2 acquisition,

to the exclusion of L1 acquisition, is a dependency on sociological factors. This includes parental discourse strategies, the status of the language in the community, availability of a speech community outside of the home, the speaker's own attitudes toward the language, and access to education in the language, among others (Montrul 2008). Although some advanced L2 speakers and heritage speakers do not qualitatively vary from adult monolingual speakers in certain grammatical domains, in general they rarely achieve full target-like proficiency in the language, instead showing effects of fossilization as well as interference from the dominant language (Montrul (2010a), Benmamoun et al. (2013b)).

2.2 WHY STUDY HERITAGE LANGUAGES?

Having considered what defines heritage speakers, we can now turn to consider why the study of heritage speakers' linguistic ability should be of any interest to the linguistic community. As discussed above, heritage speakers are a different category of language learners: their observationally incomplete knowledge of a language presents unique questions and challenges.

As discussed above, the bulk of linguistic theory has been developed based on observation of monolingual speakers, following Chomsky (1965)'s recommendations – a rare linguistic “ideal” in which a language can be examined in its most “pure” form, uncontaminated by the effects of other grammars. “Ideal” and “pure” are terms that do not indicate some real-world preference for monolinguals over multilinguals, but that have to do with how scientific study is ideally conducted – by isolating the phenomenon in question, keeping all other factors maximally constant. Analogous to isolating the variable in a scientific experiment, isolating a grammar by considering only monolingual speakers of the language is an important first step in observing and modeling properties of the language.

However, if we are considering the study of language to be the study of the human language fac-

ultly, then at some point studying only monolinguals becomes a limitation, given that most of the world is multilingual and therefore does not fit the “ideal speaker” description prescribed by early linguistic work. “Purifying the mind into a single language means destroying the actual substance we are studying – the knowledge of language in the human mind” (Cook & Newson (2007)). And indeed, linguistic theory has “reached a stage where it is possible to move further to more complex situations” (Lohndal (2013)). The field has made enough advances in studies of monolingual grammars, and we know enough (though of course not everything) about grammars of certain languages in monolingual speakers, that we can begin to explore how certain linguistic phenomena can be affected by the presence of multiple grammars in a single speaker.

In the case of heritage speakers, we are particularly interested in extending our theory to account for properties of the grammar that can be observed under conditions of decreased input and output, simultaneously under pressure from an additional dominant language. Observationally, it is clear that heritage speakers develop core aspects of their home language, but tend to simplify and over-regularize this home language, particularly in the domains of morphology and word order, as will be discussed in further detail below. All areas of language seem to be affected (vocabulary, phonology, syntax, semantics, morphology) but to various degrees depending on the phenomenon being scrutinized. From the point of view of acquisition, this raises many questions regarding the role of input. Heritage speakers therefore give us the unique opportunity to observe the effects of reduced or even interrupted input in child language acquisition, addressing the following overarching questions:

What exactly is the role of input in the development and maintenance of a language during childhood and into adulthood? When language acquisition takes place under reduced input conditions or under pressure from another language in a bilingual environment, which areas of grammar are resilient and which ones are vulnerable? What underlies the common simplification patterns observed among different heritage languages? (Benmamoun et al. (2013b))

This last question brings heritage languages into a broader linguistic debate concerning language

contact. It has been observed that the simplification patterns attested in heritage grammars often resemble the simplification patterns attested under other conditions of language contact and the emergence of new linguistic varieties such as pidgins and creoles. Echoes of the simplification in heritage languages are also observed in diachronic language change. Understanding the causal relationships between external linguistic pressures and outcomes in the grammar in heritage languages can be crucial to understanding similar processes of simplification in other types of language contact.

But in addition to addressing larger questions of language acquisition and language contact, heritage linguistics has the opportunity to inform and challenge the linguistic theory initially designed to account for monolingual speakers. Heritage languages have “the minimal scaffolding needed for a language to stand, and it has minimal design features” (Benmamoun et al. (2013b)). Benmamoun et al. (2013b) and Lohndal (2013) provide an elegant overview of this line of research and argumentation, by demonstrating that heritage speakers illustrate a loss of functional structure and that asking how and why this happens leads to powerful insights into long-standing theoretical questions in syntax, among other fields. Underlying this work is the crucial notion that at issue are not just differences in performance but also in competence. That is, the non-target-like performance exhibited by heritage speakers in natural production as well as in experimental settings is indicative of differences between heritage speakers and monolinguals in the underlying representation of the grammar.

As we will see below, the diversity in heritage speaker profiles mentioned above is reflected in a range of abilities and deficits in performance. This requires a particularly nuanced study of bilingualism, keeping track of many factors that might affect heritage speaker competence. For instance, it is important to keep in mind that not every grammatical domain and not every population will be affected by certain pressures or processes in the same way. What may be the result of attrition in one domain in one population may be the result of transfer from the dominant language in another domain (these processes will be discussed in more detail in Section 2.5). Sometimes several processes or pressures might be at play, molding the heritage grammar in intricate ways.

Outside of strictly research-based fields, research on heritage languages is motivated by a need for investigations of heritage speaker performance and competence to guide the development of pedagogical materials. As mentioned above, increasingly, heritage speakers are reporting the desire to “re-learn” their home language as a way of reconnecting with their family and culture. The presence of heritage speakers in a typical second language classroom is a challenge for all involved, as L2 learners and heritage speakers have complementary strengths and weaknesses in the classroom. Instructors must cater to two very different demographics in their classroom. The L2 peers themselves are adept at pencil-and-paper exercises but are often intimidated by the relative comfort with which the heritage speakers – typically more comfortable speaking than writing in their HL – speak up in class. Meanwhile the heritage speakers themselves do not always find the methods used in L2 classrooms to be effective at improving their own abilities and literacy. In some parts of the country, particularly the West Coast, efforts have been taken at the high school and college level to develop specialized language classes for heritage speakers, most often of Spanish and Chinese. However, even in these cases the pedagogical methods implemented in heritage classrooms are largely based on teachers’ personal experiences and trial-and-error rather than on tools and materials based on linguistic research on heritage languages. Linguistic research needs to identify what heritage speakers know (and what psycholinguistic processes are involved), in order to experimentally determine whether heritage speaker performance in the HL can be substantially and permanently affected by classroom intervention.

To date there is relatively little of this systematic and theoretically driven research on HLs, although the trends in language classrooms combined with the recognition of potential implications of heritage language research on current linguistic theory are driving a positive change. The remainder of this chapter is dedicated to providing an overview of the research that has been done on heritage languages, with particular emphasis on work on grammatical gender, but it should be noted here that there is a substantial gap in the literature. Heritage linguistics largely lacks studies

that systematically compare groups of heritage speakers to each other. Comparisons across studies are difficult to make due to varied experimental methods and selection criteria for heritage speaker participants. In particular, comparing heritage speakers of the same language but with different dominant languages stands to give substantive insight into the role of transfer (from the dominant language) in the development of heritage languages, but most work has been done with English as the dominant language. In the only study of such scope, Kim (2007) compared the processing of reflexives between heritage speakers of Korean living in the United States with those living in China; Kim showed that properties of the dominant language did not in fact play a role in how the heritage speakers processed reflexives. After providing an overview of previous work on heritage languages, this chapter will outline how the present project will contribute to our understanding of grammatical gender by comparing several groups of heritage speakers in a manner similar to that implemented in Kim (2007).

2.3 PREVIOUS WORK ON HERITAGE LANGUAGES

Heritage linguistics has noted that all parts of the heritage grammar differ from the monolingual grammar (at least observationally, at the surface level), and to varying degrees. Work on heritage linguistics hails from many subfields of linguistics. The studies presented in this dissertation are pertinent only to the morphosyntax of heritage languages (Section 2.3.4), but the remainder of this section is intended to provide a broad overview of significant findings from heritage linguistics. It is by no means exhaustive, but provides an important insight into the variability in heritage grammars. As this survey will show, in some domains of grammar heritage speakers differ notably from the monolingual baseline, whereas in others their performance appears to be only minimally affected. Even further, in some instances there appear to be asymmetries between languages wherein a particular grammatical domain is affected in heritage speakers of a certain language but not in heritage

speakers of a different language. For detailed discussion, I refer the reader to the papers discussed here and the citations therein.

2.3.1 VOCABULARY

The lexicon of a heritage speaker is very often a direct reflection of the restricted nature of heritage language acquisition. Recall from the earlier discussion that use of the heritage language is typically limited to the context of the home, and depending on properties of the community, church or community centers. The dominant language is used in all other contexts, including academic and professional ones. It is rather unsurprising then, that heritage speakers are most comfortable with vocabulary related to common objects used in the home and other vocabulary from their childhood.

As demonstrated by Polinsky (1997, 2006), lexical proficiency is correlated with structural accuracy in heritage speakers of Russian. When asked to identify words from a list of 200 common lexical items, those heritage speakers who knew a greater proportion of the list also had better control of agreement, case, and subordination in spontaneous speech. As mentioned above, these results were replicated for heritage speakers of Arabic, Russian, Polish, Armenian, Korean, and Lithuanian (see Godson (2003), Albirini et al. (2013), a.o.).

There are also certain asymmetries in the lexical proficiency of heritage speakers. Polinsky (2005) found that low proficiency heritage speakers of Russian have better control of verbs than of nouns and adjectives. These findings were replicated for heritage speakers of Korean by Lee et al. (2012). From the point of view of external pressures on the language, this result is unsurprising when taking into consideration the fact that nouns are more frequently used in code-switching than verbs are (Poplack (1980)). From a theoretical perspective, Polinsky (2005) argues that loss of nouns is simply less costly to the heritage speaker than the loss of verbs. This is because verbs carry relational information and are therefore semantically more dense, whereas nouns carry less relational information, and their loss is not as significant to the remainder of the grammar.

2.3.2 PHONETICS AND PHONOLOGY

Among the other parts of the grammar, the phonology of heritage speakers is usually quite good. This is supported by ample evidence in which heritage speakers and L2 speakers of similar morphosyntactic proficiency were directly compared to each other in their ability to produce utterances in the target language with monolingual-like phonology [Au et al. \(2002\)](#), [Oh et al. \(2003\)](#), [Yeni-Komshian et al. \(2000\)](#), [Knightly et al. \(2003\)](#). Similarly, while L2 speakers struggle to distinguish between various phonemes that are not a part of their native language, heritage speakers appear to be very adept at phoneme differentiation, including in tonal languages ([Chang & Yao \(2016, 2019\)](#)). [Oh et al. \(2003\)](#) demonstrate that the phoneme perception of heritage speakers of Korean appears to be largely unimpeded, although [Ahn et al. \(2017\)](#) suggest that heritage speakers for whom the onset of bilingualism occurs earlier are less likely to have this advantage over L2s.

This is not to say that heritage speakers do not differ from the baseline. While they show a significant advantage over L2s in producing target-like phonology, heritage speakers do display some nonnative phonological features. These differences between heritage speakers and baseline monolinguals are measurable and systematic, and allow some insight into the source of the discrepancy. [Godson \(2004\)](#) found that heritage speakers of Western Armenian had nonnative vowel phonology that showed transfer effects from the English vowel system.

Some research pursues the question of whether heritage phonology can be influenced to become even more target-like. [Fuchs et al. \(under revision\)](#) observed the phonology of heritage speakers before and after a semester-long Spanish course designed specifically for heritage speakers. Both before and after the semester, student participants were asked to record a narrative of over 50 phrases. Measures of voice onset time (VOT) and vowel formants showed that overall, while their phonetics was not perfectly target-like, it did become closer to the monolingual baseline after the semester of instruction in Spanish.

2.3.3 SEMANTICS

Investigations of the semantics of heritage languages are less commonly undertaken, but research has demonstrated that this domain of the heritage grammar is also affected. One common area of interest is the semantics of articles. To illustrate, [Ionin & Montrul \(2010\)](#) and [Montrul & Ionin \(2012\)](#) found that heritage speakers of Spanish living in the United States have a strong tendency toward using bare nouns with generic reference in subject position – a construction that is disallowed by the monolingual grammar, but appears to be the result of transfer from English, which allows generic bare nouns. Studies of heritage semantics have also been interested in semantically based inherent case, such as the genitive of negation in Russian ([Polinsky \(1997, 2006\)](#)). Similarly, heritage speakers of Spanish struggle with differential object marking (DOM) ([Montrul \(2004\)](#), [Montrul & Bowles \(2009\)](#)), mentioned again in the morphosyntax discussion below. More recently, [Scontras et al. \(2017\)](#) found that heritage speakers of Chinese tend to restrict themselves to surface-scope readings of sentences with multiple quantifiers for which monolingual speakers of Chinese allow or prefer an inverse-scope interpretation.

2.3.4 (MORPHO)SYNTAX

In the subfield of syntax, one of the primary observations about heritage languages is that they tend to impose strict word order where the monolingual grammar allows flexibility. The well documented lack of sensitivity to inflectional morphology among heritage speakers results in their relative inability to use case markers as cues to syntactic position, and they must use word-order information to determine argument structure. For heritage speakers of Spanish and Russian this means relying on SVO order, and for Korean speakers this means relying on SOV word order ([Song et al. \(1997\)](#)). As an example, heritage speakers of Spanish had trouble with comprehending sentences with preverbal objects ([Montrul \(2010b\)](#)). The effect can be seen also in relative clauses: heritage

speakers of Russian have difficulty with comprehension of object relative clauses, which have a non-standard word order, while they have no trouble understanding subject relative clauses, which maintain SVO order (Polinsky (2008b), O’Grady et al. (2001)).

Strict adherence to word order plays another role in heritage Spanish. Albirini et al. (2011) found that heritage speakers of Spanish allow closest conjunct agreement, something that is ungrammatical in the monolingual baseline. As Albrini et al. argue, this seems to suggest that heritage speakers use adjacency to compute agreement with a coordinate noun phrase subject. In the baseline grammar, agreement with a coordinated noun phrase is determined by a set of complex interface constraints, but Albrini et al. propose that heritage speakers favor a much simpler computation based on linear proximity or adjacency.

More broadly, there is a general trend in heritage linguistics findings that suggests that phenomena that require incorporating external knowledge to determine the contextual appropriateness of an utterance are particularly vulnerable in heritage grammars (the Interface Hypothesis, in the updated sense of Sorace (2011)). These phenomena “require access to two separate systems with overlapping but convergent primitives and principles and rule of combination [...] they require knowledge of how to map one component onto the other, no simple task in any framework” (Benmamoun et al. 2013b:54).

A well-studied phenomenon that supports this proposal is the licensing of null and overt subjects in heritage grammars. Licensing of null subjects occurs at the syntax-discourse interface and is particularly affected in heritage grammars (Sorace (2004)): heritage speakers tend to overuse overt subjects where null subjects are more pragmatically appropriate in the monolingual baseline. Heritage speakers of Spanish and Russian overuse overt subjects in topic shift and switch reference contexts (Montrul (2004), Polinsky (2006), Silva-Corvalán (1994)). Similar results have been found for Hungarian (De Groot (2005)), Hindi (Mahajan (2009)), Tamil and Kabardian (Polinsky (1997)), Spanish (Silva-Corvalán (1994), Montrul (2004)), Polish (Polinsky (2006)), and Arabic (Albirini et al. (2011)).

Studies of heritage syntax are an excellent case in point for the discussion above, that a result in a single domain of the grammar should not be extrapolated to other parts of the domain. Certain parts of heritage syntax are indeed very vulnerable, including word order, and long-distance dependencies (Montrul (2010a), Benmamoun et al. (2013b), Polinsky (2018)), but this is not the case for all areas of syntax. For instance, Montrul (2005) that even low proficiency speakers of Spanish understand unaccusativity – a surprising result given challenges with many other syntactic structures. It is not even the case that a certain domain of syntax is equally affected in all heritage languages. As mentioned above, heritage speakers of Russian have difficulty in comprehending non-subject relative clauses (Polinsky (2008b)), but heritage speakers of Spanish have no trouble comprehending various relative clauses (Sánchez-Walker (2012)). While imposing strict word order is a hallmark of many heritage languages, per the discussion above, heritage speakers of Spanish have target-like or almost target-like control of V2 word order, which even highly proficient L2 learners struggle with (Håkansson (1995)).

One of the hallmarks of heritage languages is an abundance of errors in agreement and case-marking, which has been well documented for a range of grammars and which this dissertation is primarily engaged in. These errors are more pronounced in the nominal domain – where the focus has largely been on gender marking and case marking – than in the verbal domain (Benmamoun et al. (2013b), Bolonyai (2007)). For instance, Hindi heritage speakers used non-target case marking at a rate of 23-27%, but their error rate for verbal morphology remained under 7% (Montrul et al. (2012)). Similarly, Polinsky (2006) shows that Russian heritage speakers make morphosyntax errors in the nominal domain at a rate of 40%, but in the verbal domain their error rate does not exceed 20%. Heritage speakers of Egyptian and Palestinian Arabic have target-like nominal concord about 64% of the time, but in the verbal domain they are target-like about 83% of the time (Albirini et al. (2011), followed up experimentally by Albirini et al. (2013)). A similar asymmetry has been found for heritage speakers of Hungarian (Fenyvesi (2000), De Groot (2005), Bolonyai (2007)). Such an

asymmetry between agreement in the nominal and verbal domains seems to suggest that the process responsible for agreement in the two domains are not the same – such a proposal based on heritage language facts would be in line with theoretical arguments made by Norris (2014) in favor of analyzing nominal concord as a process distinct from verbal agreement (Benmamoun et al. (2013b)).

Various parts of verbal inflection have been isolated and observed in studies of heritage speakers' command of morphosyntax. Primary among them is of course subject-verb agreement (see Moag (1995) for work on heritage speakers of Asian languages). In the realm of tense, we find yet another asymmetry: heritage speakers of Russian control regular forms of the past tense and present tense, but they confuse the perfective and imperfective (Montrul (2002), Polinsky (2006), Silva-Corvalán (1994)). In Spanish, heritage speakers particularly struggle with their command of the subjunctive mood (Montrul (2007), Silva-Corvalán (1994)) and conditional (Silva-Corvalán (1994)). Rothman (2007) also shows that heritage speakers of Brazilian Portuguese do not develop inflected infinities, which Rothman attributes to the fact that inflected infinitives belong to written registers typically acquired in school, which heritage speakers do not attend in their heritage language.

Back in the nominal domain, heritage speakers notoriously struggle with production and comprehension of case. Polinsky (2006, 2008c) shows that heritage speakers of Russian have a reduced case system: where monolingual Russian has six cases, low proficiency heritage speakers of Russian have reduced this case system to just two cases (the nominative and accusative). Song et al. (1997) found a similar reduction of the case system in heritage Korean. Heritage speakers of Spanish omit the preposition *a* that is obligatorily used with dative experiencer subjects and animate direct objects in the monolingual grammar Montrul & Bowles (2009). This omission of case has also been observed for heritage speakers of South Asian languages (Moag (1995)) and is often tied to a more pronounced reliance on word order to determine argument structure.

As for agreement in the nominal domain, research in this vein is most often concerned with errors in plural agreement and gender agreement. For instance, Benmamoun et al. (2008) show that

heritage speakers of Arabic make errors in plural agreement at a rate of about 30%. Benmamoun et al. note that the errors observed in the heritage grammar are the same made by monolingual children during acquisition: both speaker groups overextend use of the feminine suffix -aat to masculine contexts. Similarly for Greek, Zombolou (2011) found that heritage speakers of Greek living in Argentina overextended the default plural for each of the three Greek genders, similar to the overextension errors made by preschool-age monolingual Greek children acquiring the language.

Errors in gender assignment and agreement are similarly – if not more – common to heritage languages, and have been observed in languages like Russian, Spanish, and Swedish (Polinsky (2008c), Montrul et al. (2008), Håkansson (1995)). Typically, monolingual children of Russian and Spanish have full control of gender by age 4, but heritage speakers of these languages show an error rate between 5% and 25% (Polinsky (2008c), Montrul et al. (2008)). In fact, Polinsky (2008c) shows that low proficiency heritage speakers of Russian reduce the gender system: monolingual Russian has masculine, feminine, and neuter, but the heritage grammar has only masculine and feminine genders. In light of observed differences in gender assignment and agreement among heritage speakers, this dissertation asks whether and how heritage speakers can use gender information in online tasks. A more thorough review of recent work on gender in Heritage Spanish and Heritage Polish will be undertaken in Chapter 3 and Chapter 4, respectively.

2.4 THE SIGNIFICANCE OF TASK TYPE

An important lens through which to consider the body of literature summarized above is the now commonly-accepted observation that heritage speaker performance on a linguistic task is very much dependent on the modality of that task (Montrul (2010a), Montrul & Ionin (2012), Montrul et al. (2014)). For instance, Montrul et al. (2008) found that heritage speakers of Spanish performed better than L2s at an oral picture naming task, but performed worse than the L2s at a written recogni-

tion task as well as a written comprehension task. Montrul et al. (2014) performed a direct comparison of heritage speakers with L2s and baseline monolinguals, with similar results: heritage speaker performance was more target-like than that of L2s in a word repetition task that tested more implicit linguistic knowledge, but more like L2s in the two tasks that tested speaker knowledge more explicitly (through grammaticality judgment and gender monitoring tasks).

A key factor affecting heritage speaker performance appears to be whether the task requires metalinguistic competence and awareness. Heritage speakers typically learn the HL at home but rarely receive formal schooling in the language, which severely limits the amount of metalinguistic awareness they may have about their HL. The input to their grammar has been entirely naturalistic and aural, rather than formal or presented in a written medium with access to written registers of the language. By contrast, L2 learners of a language receive input to the grammar primarily in written form – in the classroom, through textbooks and written exercises. As Montrul & Ionin (2012), Montrul et al. (2014) demonstrate, this results in differing advantages for the two speaker groups. Written versus oral tasks tap into different implicit or explicit knowledge (Ellis (2005), Ellis et al. (2006)). Accordingly, heritage speakers perform better at those tasks that probe implicit linguistic knowledge, and as aural learners, heritage speakers fair better in oral production and aural comprehension tasks, but underperform on written tasks. Additionally, it need not be the case that heritage speakers perform similarly in production versus comprehension of a particular phenomenon. Thus, much work compares these two domains, and many experimental designs incorporate some type of elicited production task in combination with a comprehension task. For a detailed discussion of how heritage speaker comprehension and production compare across subfields (syntax, phonetics, morphology, etc) see Polinsky (2018).

The takeaway from research into task modality is that care needs to be taken when designing experimental tasks for heritage speakers to ensure that the modality of the task will allow for an accurate assessment of the heritage speakers' knowledge of a certain part of the grammar. Studies that

aim to determine how target-like heritage speakers might perform in a certain area should select an appropriate experimental design; a written task that requires metalinguistic awareness might put heritage speakers at a disadvantage, leading to an underestimate of the speakers' knowledge and ability. Tasks that tap into implicit linguistic knowledge through investigating aural or visual comprehension give heritage speakers a boost and are therefore more likely to be an accurate indicator of heritage speaker capabilities.

2.5 DEVELOPMENTAL TRAJECTORIES

An open question in heritage language research is the developmental trajectory of heritage grammars. While at the outset heritage grammars often has a similar starting point as monolingual grammars, they ultimately diverge from the monolingual grammar, at least on the surface. What process(es) intervene to yield the heritage grammar? Research has presented several candidates, which will be introduced below. The studies presented in this dissertation will not bear in any way on the hypotheses presented in this section. However, in recognition of the fact that heritage grammars begin to diverge from monolingual grammars around the onset of formal schooling, Chapters 3 and Chapters 4 will provide an overview of the literature on gender acquisition in Spanish and Polish, respectively.

As discussed, the differences between monolingual grammars and corresponding heritage grammars are striking and well documented, with consistent patterns in the structures and processes that are most non-target-like. There are several candidate forces that may influence and morph a particular part of the heritage grammar: divergent acquisition, attrition, transfer, incipient changes in the input, and a preference for default-like structures. It is often assumed that more than one of these may be at play in any given heritage language or even any subpart of the heritage grammar. Here I will provide an overview of the main developmental trajectories studied by work on heritage linguistics.

tics.

Attrition is the notion that child heritage language speakers can be quite proficient in their language but then somehow “lose” their ability over time. More formally, attrition implies that a given grammatical structure was fully acquired and mastered by the child heritage speaker, before weakening or being lost entirely as a result of an extended period of time with reduced input and output. [Seliger \(1996\)](#) provides the most commonly cited metric for identifying attrition: “temporary or permanent loss of language ability as reflected in a speaker’s performance or in his or her inability to make grammaticality judgments that would be consistent with native speaker monolinguals of the same age and stage of language development.” While attrition is commonly discussed in the context of heritage speakers, it should be noted that instances of loss of linguistic proficiency among first generation immigrants also fit the description. In second-generation immigrants, the extent of attrition may be related to the age of onset of bilingualism: children who are exposed to the L2 before puberty lose proficiency in their L1 much quicker and to a greater extent than those who are exposed to the L2 after puberty or as adults ([Bylund \(2009\)](#), [Pallier \(2007\)](#), [Ammerlaan \(1996\)](#), [Hulsen \(2000\)](#), [Montrul \(2008\)](#)). In extreme cases of completely interrupted input before puberty, children may experience total language loss of the L1, as has been documented for international adoptees.

As mentioned, not all parts of the heritage grammar need be impacted by attrition. One alternative trajectory of development is divergent acquisition. This is the notion that certain developmental delays that start in childhood are never overcome. This can be illustrated for sequential bilinguals: at the moment the monolingual soon-to-be-heritage child speaker enters school, she starts to diverge from her monolingual counterparts in amount of input and language use in the L1. At this point, the likelihood that she will fully acquire and master grammatical structures that are learned after age 5 or in written registers taught in schools becomes much lower. Such a developmental trajectory has been observed for the Spanish subjunctive. [Blake \(1983\)](#) shows that monolingual Spanish children do not master the subjunctive until after age 10. The fact that heritage speakers struggle with

use of subjunctive is therefore likely the outcome of never fully acquiring the subjunctive due to restricted Spanish-language input before age 10 (Martinez-Mira (2009), Montrul (2009), Potowski et al. (2009), Silva-Corvalán (1994)). We can tease apart divergent acquisition from attrition through longitudinal studies of heritage speakers (Anderson (1999), Merino (1983), Silva-Corvalán (2003)) or by comparing child heritage speakers with adult heritage speakers of the same language (see Polinsky (2008b) for a study on Russian relative clauses).

The development of a heritage grammar may also be influenced by external pressures, such as transfer – how the heritage speaker’s dominant L2 might affect the grammatical structures of the heritage L1. Transfer effects have been well documented for adult L2 acquisition, where the native L1 impacts the acquisition of the new L2 (Odlin (1989), White (1989), Gass & Selinker (1992), Schwartz & Sprouse (1996), Jarvis (1998)). Somewhat counterintuitively, it is also possible for the second language to affect the grammatical structures of the L1; this has been studied particularly in L2 acquisition (Cook (2003), Pavlenko & Jarvis (2002), Seliger (1996)). Transfer can also be seen in other instances of language contact, such as creoles and pidgins. Effects of transfer can be identified by comparing the effects of different dominant languages on heritage speakers of the same language (as in Kim (2007), wherein heritage speakers of Korean living in the United States were compared to heritage speakers of Korean living in China) or by testing heritage speakers whose majority language is close to the HL in complexity of the relevant grammatical structure. While this dissertation does not test the effects of transfer on the use of gender in online tasks, Section 2.6 discusses how the study has been extended to do so, pending data analysis on a study conducted in Germany.

Importantly, it is not sufficient to compare heritage speakers with the baseline to identify transfer. The language of the first generation is the input to the language of the second-generation heritage speakers; therefore, if the grammar of the first-generation immigrants has already changed and differs from the grammar of monolingual speakers in the home country, then it is possible that the heritage language property in question is simply a matter of having acquired that non-target-like va-

riety from input. This is because some properties of the heritage language are the result of incipient changes in the grammar of the first-generation immigrants. This in turn touches on the question of whether immigrant communities speak different varieties of a language than do monolingual speakers in the country of origin. At least differential object marking in Spanish, the answer appears to be “yes”. Montrul (2004) and Montrul & Bowles (2009) found that heritage speakers of Spanish omit DOM, but a follow-up in Montrul & Sánchez-Walker (2013) found that first generation immigrants do so as well. The implication is that heritage speakers of Spanish do not acquire monolingual-like DOM because the input to their grammar already lacks target-like DOM.

One final developmental trajectory common to heritage languages is a preference for default-like structures. Under the pressure of speaking a non-dominant language, some grammatical structures are recast to become more user-friendly. For instance, heritage languages often eliminate irregular morphology, simultaneously giving rise to analyticity (Benmamoun et al. (2013b,a)). Similarly, heritage languages often impose rigid word order (Isurin & Ivanova-Sullivan (2008), Ivanova-Sullivan (2014)) and impose placement of closely associated items next to each other, in keeping with Behaghel’s First Law (Behaghel (1909), Haiman (1983)). Heritage speakers show a preference for dependencies that target only the highest structural constituent; Polinsky (2008b) shows that heritage speakers of Russian can comprehend subject relative clauses, to the exclusion of all other types of relative clauses. Heritage languages also often show an absence of nesting dependencies (Benmamoun et al. (2013b,a)) and a lack of non-compositional structures (Dubinina (2012), Rakhilina & Marushkina (2014)).

Again, at present this study does not attempt to address the question of which process(es) affect the development of grammatical gender or its use in online tasks in the heritage grammar. Nevertheless, awareness that certain parts of the grammar may or may not be fully acquired at some point during heritage language acquisition is important to keep in mind, motivating an overview of the literature on acquisition of gender in Spanish (Chapter 3) and Polish (Chapter 4).

2.6 CONTRIBUTION OF THIS PROJECT

As discussed in the Introduction, this project brings together a theoretical understanding of gender as a syntactic feature with evidence from monolingual and heritage acquisition to investigate gender in the adult heritage grammars of Spanish and Polish. Spanish provides an ideal starting point given the relative wealth of previous work on the subject. One of the main contributions of the study on Spanish is, informed by investigations of how task modality can impact performance, employing a very sensitive implicit methodology (*eye-tracking*) to measure linguistic ability. In fact, much of the experimental design and choices for experimental procedure is motivated by the discussion in this chapter pertaining to what experimental contexts allow heritage speakers to perform in a way that is most representative of their linguistic knowledge and ability.

The subsequent study on Polish is motivated by the observation that crosslinguistically gender can be instantiated as more complex than a binary feature. Exploring the three-way gender distinction in Polish allows us to take a step in the direction of a more generalizable understanding of how heritage speakers use gender in online processing. The study on Polish also constitutes, to my knowledge, the first psycholinguistic investigation of gender in Polish, contributing to the literature by proposing a hierarchical relationship between the gender feature values in the language.

Further, a comparison across the studies suggests an asymmetry between how gender information on articles and modifiers are accessed in online tasks. The unexpected finding is that adjectives appear to be easier for heritage speakers than articles are, a distinction that is discussed in detail in Chapter 5. I ultimately leave open the question of what is at the root of this contrast, but several considerations point to a rich field for further investigation.

3

Eye-tracking study: Heritage Spanish

SPANISH IS A GREAT STARTING POINT for an investigation of how heritage speakers use their knowledge of gender in online processing. Gender information is transparent on determiners, nouns, and adjectives in Spanish. As will become clear, there is an abundance of literature on the grammatical gender feature in Spanish, both from the formal perspective and the acquisitional perspective, and the use of grammatical gender in online processing has already been investigated

among monolingual children and adults, as well as L2 speakers of Spanish. Interpreting the results of the study in this context allows us to draw broader implications regarding the language processing abilities of heritage speakers.

This chapter will begin by providing the relevant information on the surface facts of gender in Spanish (Section 3.1.1) and the research that has been done on gender in the domain of monolingual and bilingual acquisition (Section 3.1.2). The present study will be motivated in more detail through a discussion on previous eye-tracking studies of anticipatory use of grammatical gender in Spanish (Section 3.1.4). The experimental design and methodology will be presented in Section 3.2, with predictions and details regarding participants laid out in Section 3.3 and Section 3.4, respectively. The results of the study are presented in Section 3.7, with discussion of the implications in Section 3.8. In Section 3.9 I outline outstanding questions and next steps, motivating the study on Heritage Polish that is presented in Chapter 4.

3.1 BACKGROUND

3.1.1 GENDER IN SPANISH

As touched on in previous chapters, Spanish has two surface genders: masculine and feminine. These are distributed roughly equally in the lexicon, with masculine nouns composing 52% of the lexicon and feminine nouns composing 45% (Bull (1965)).* As discussed in Chapter 1, the most reliable indicator of the gender category of a noun is the form of associated words in the nominal domain. In Spanish, these associated words include articles, demonstratives, and adjectives; outside of the nominal domain, predicate adjectives must also inflect for gender. (1) illustrates the form of associated words (in this case, the definite article and an adjective) of a masculine noun, while (2) illustrates the form of associated words of a feminine noun.

*The remaining 3% of the lexicon consists of epicene – or gender-neutral – nouns.

- (1) a. \underline{el} libro $\underline{roj-o}$
 the.M book.M red-M
 ‘the red book’
- b. \underline{la} manzana $\underline{roj-a}$
 the.F apple.F red-F
 ‘the red apple’

As these examples show, the definite article varies in form depending on the gender of the head noun – *el* is the form associated with the masculine, and *la* is the form associated with the feminine. From the examples above it is also evident that adjectives inflect to match the gender feature of the head noun (more on this below).

As previously discussed, the morpho(phono)logical form of a word is rarely a fully reliable indicator of gender category, but in Spanish there are certainly some correlations between gender and what is referred to as *word markers* (Harris (1985, 1991)). Chief among them is the association between the word marker *-o* and the masculine gender, as well as the word marker *-a* and the feminine gender. (2) illustrates a few examples of this correlation both in human/animate noun pairings (2-a) as well as in inanimate nouns (2-b).

- (2) a. *el hijo* ‘the son’
la hija ‘the daughter’
el camarero ‘the waiter, masc.’
la camarera ‘the waiter, fem.’
- b. *el libro* ‘the book, masc.’
el gancho ‘the hook, masc.’
la mesa ‘the table, fem.’
la manzana ‘the apple, fem.’

The relationship between the canonical *-o/-a* word markers and masculine/feminine gender is

certainly not a one-to-one correspondence. The examples in (3) illustrate that there are other possible noun endings that can belong to either gender (with varying frequencies), and in (4) we see that nouns with word markers *-a* can even belong to the masculine gender, just as nouns with word markers *-o* can belong to the feminine gender.[†]

- (3) a. *el sol* ‘the sun, masc.’
 el arroz ‘the rice, masc.’
 la flor ‘the flower, fem.’
 la sal ‘the salt, fem.’
- b. *el mapa* ‘the map, masc.’
 el día ‘the day, masc.’
 la mano ‘the hand, fem.’
 la foto ‘the picture, fem.’

MASCULINE AS THE DEFAULT GENDER

Considering the two possible features within the gender category, Harris (1991) argues that masculine is the default gender, following Prado (1982) and Roca (1989) – that is, per the discussion in Section 1.3, what is recognized as the masculine gender is in fact “the absence of any information about gender in lexical entries” (Harris 1991:44).[‡] Evidence for masculine as the default gender (otherwise also referred to as the unmarked or unspecified gender), comes in various forms. Harris (1991) cites evidence from metalinguistic use of non-nominal words as nouns. In (4), the preposition *para*

[†]Harris (1991) notes an asymmetry here, in that there are more masculine nouns with the *-a* word marker than feminine nouns with the *-o* word marker.

[‡]This is in contrast with his earlier work, in which he claimed gender in Spanish is binary, represented as [+fem] for feminine, [-fem] for masculine Harris (1985). Arguments for this binary analysis of Spanish gender center on the fact that *el* and *la* appear to be equally specified morphologically, as they are also for *unos/unas* and *los/las*. In fact the only instance in which the masculine morphology is subtractive on Spanish determiners is for the singular indefinite articles *un/una* (Dominguez et al. (1999), Alarcón (2006)).

‘for’ is used metalinguistically as a noun. As a preposition, *para* is inherently genderless, but to be used in the syntactic position in which it appears in (4), it must be assigned a gender. As we see from the inflection on *demasiados*, it is clearly treated as a masculine noun, suggesting masculine gender is assigned when other gender assignment rules fail.

- (4) Tienes demasiados “paras” en ese párrafo.
 Have.2SG too.many.M.PL “para”.PL in this.M paragraph.M
 ‘You have too many “paras” in this paragraph.’

In a similar vein, [Roca \(1989\)](#) offers examples of default masculine agreement when the gender of the referent is unimportant or unknown. In (5), the subject is the noun *nadie* ‘nobody’, and the predicate *mal-o* ‘evil’ is marked for masculine agreement.

- (5) Para él, nadie es mal-o
 For him, nobody be.3SG evil-M
 ‘For him, nobody is evil.’

Default agreement also arises in Spanish in coordinated NPs, in which the two coordinated nouns are of different genders, as in (6) (drawn from [Fuchs et al. \(2015\)](#)). As this example demonstrates, in such cases the inflection on agreeing predicates indicates default masculine gender agreement.

- (6) El libro y la pintura son caros /*caras.
 the.M book.M and the.F painting.F be.3PL expensive.M.PL /expensive.F.PL
 ‘The book and the painting are expensive.’

Taken together, evidence from neologisms and unknown referents constitutes a convincing and widely accepted argument that gender is a privative feature in Spanish, with the masculine as the default gender feature. One way of representing this privative opposition is to follow [Harris \(1991\)](#) and [Fuchs et al. \(2015\)](#), among others, in representing the feminine gender as the specified [fem], and the masculine feature as the lack of a gender specification []. [Kramer \(2015\)](#) distinguishes between

$n\ i[+FEM]$	interpretable feminine	<i>el niño</i> ‘the boy’
$n\ i[-FEM]$	interpretable masculine	<i>la niña</i> ‘the girl’
$n\ u[+FEM]$	uninterpretable feminine	<i>la mesa</i> ‘the table’
n	(default) uninterpretable masculine	<i>el libro</i> ‘the book’

Table 3.1: Gender features in Spanish (Kramer (2015))

notional and grammatical gender and provides a more detailed breakdown of gender in Spanish, positing four possible feature values, provided in Table 3.1.

GENDER ON ADJECTIVES

As mentioned above, the gender feature on the head noun also determines gender agreement throughout the nominal phrase. The fact that determiners inflect for gender was discussed above; adjectives in Spanish typically also inflect to match the gender of the noun they modify, with the *-a* marker again indicating feminine, and *-o* indicating masculine gender agreement, as in (7).

- (7) a. *el libro rojo* ‘the red book’
b. *la manzana roja* ‘the red apple’

There are, however, adjectives that do not inflect for gender, but instead keep the same form regardless of the gender of the head noun, as illustrated for *grande* ‘big’ in (8). The group of adjectives that follow this pattern includes common adjectives such as *triste* ‘sad’, *interesante* ‘interesting’, and *azul* ‘blue’, among others.

- (8) a. *el libro grande* ‘the big book’
b. *la manzana grande* ‘the big apple’

So far we have only seen adjectives occurring post-nominally, and indeed, [DET + N + ADJ] is the typical word order within the Spanish nominal phrase. Some adjectives show flexibility, in that they

can occur either prenominally or post-nominally, although this is often associated with a change in meaning, as illustrated in (9). Other elements internal to the nominal phrase that can appear before the noun are quantifiers.

- (9) a. *mi amiga vieja* ‘my old friend (friend is old in age)’
b. *mi vieja amiga* ‘my old friend (they have been friends for a long time)’

Finally, adjectives in the Spanish nominal phrase can also appear in structures of the form DET + ADJ, with no overt noun, as in (10). In these cases of nominal ellipsis, the referent is identified using gender and number features on the determiner and adjective.

- (10) a. *la roja* ‘the red (one), fem.’
b. *los tontos* ‘the silly (ones), masc.’

With this understanding of the relevant aspects of Spanish nominal syntax, we can now turn to consider how children acquire the gender feature and gender agreement within the nominal phrase during L1 and bilingual language acquisition (Section 3.1.2) and how this knowledge is deployed by adult, child, and L2 speakers of Spanish (Section 3.1.3).

3.1.2 GENDER IN ACQUISITION

Because the status of a particular domain of the heritage grammar may be tied to the timing and nature of acquisition, I provide here a survey of work on the acquisition of gender in monolingual and bilingual child acquisition before turning to previous work on gender in adult heritage speakers.

MONOLINGUAL ACQUISITION

Longitudinal studies of monolingual L1 acquisition of gender in Spanish suggest that the gender system has reached pretty much complete acquisition, at least on determiners, around age 2;7 to

3 (Soler (1984), Hernandez Pina (1984), Pérez-Pereira (1991), Mariscal (1996, 2001), López-Ornat (1997), Lleó (1997)). By this age, children's production of determiners is largely error free, with less agreement errors in the nominal domain than in the verbal domain (López-Ornat et al. (1994)). When children do make errors, they change the ending of the noun to better match the canonical endings in the gender category. For example, Clark (1986) reports children saying *la mujala* (intended: *la mujer* 'woman, fem.') and Hernandez Pina (1984) documents children saying *mota rota* (intended: *moto rota* 'broken motorbike'). Spanish children are said to acquire the gender system relatively quickly, but they do struggle with nouns that end in *-e*, which can belong to either the masculine or the feminine gender. In a study on gender assignment of 160 monolingual Spanish-speaking children, Pérez-Pereira (1991) found that children are more likely to rely on formal properties of a word (i.e. morphophonological properties) than on extralinguistic ones (in this case, natural gender) in assigning gender to an unfamiliar word.

Notably, even before acquisition of gender marking on determiners, monolingual children appear to be aware of the presence of a determiner-like prenominal element. López-Ornat (1997) notes that very young children in this linguistic group produce nouns with a prenominal vowel (*e pie* 'the foot', *a queca* 'a doll', *u fo* 'a flower'; López-Ornat (1997)). Crucially, these prenominal vowels correspond to the vowel of gender-marked determiners (*el, la, un, una*). Lleó (1998) proposes that these are a kind of "protodeterminer", and that, with experience, monolingual and bilingual children learn to segment these protodeterminer+noun chunks into a determiner and noun. But this initial stage of an unanalyzed chunk suggests that there is a very tight link between the gender-marked determiner and the noun in Spanish, a notion that will be relevant in the discussion of eye-tracking studies below.

There is some evidence that agreement with determiners is mastered before agreement with adjectives (Pérez-Pereira (1991)), although there is considerably less data on this matter. Mariscal (2009) combined observational data with experimental data to supplement previous findings, and con-

firmed that error-free production of gender on determiners is observed before error-free production of gender agreement on adjectives. Where children make errors in adjective agreement, the strong tendency is to overuse masculine agreement. Adult-like mastery of nominal ellipsis appears to be attained around age 3-4 (Snyder et al. (2001)).

The delay in acquisition of gender agreement on adjectives may in part be due to frequency effects. According to the Diccionario de Frecuencias, the feminine article *la* is the second most frequent word in Spanish, while the masculine article *el* is the fifth most frequent word in the language (Alameda & Cuetos (1995)). In child-directed Spanish, both articles appear among the top ten most frequently used words, as determined by the longitudinal corpus María in CHILDES (López-Ornat et al. (1994)). Comparatively, adjectives are much less frequent in the lexicon.

BILINGUAL ACQUISITION

Bilingual acquisition of gender in Spanish has been much more studied. For the purposes of heritage linguistics broadly, investigating to what extent bilingual children are proficient in gender agreement can begin to answer the question of what developmental trajectory leads to the gender performance and competence observed and analyzed for adult heritage speakers. The main questions here are: Do bilingual children ever fully master gender (and gender agreement) in Spanish? And if so, is this competence subject to attrition? For the purposes of this dissertation, the question is whether the phenomenon of interest – gender agreement – is acquired by bilingual children in a monolingual-like manner. Many of the studies on heritage speaker children focused on a comparison between heritage speakers and English-dominant L2 learners who began to acquire Spanish after puberty, but the discussion here will focus on the results reported for heritage speakers.

The first work in this vein is credited to a cross-sectional study by Sanchez-Sadek & Others (1975), who tested bilingual children in a school in Los Angeles. The researchers tested children in kindergarten, 1st grade, 2nd grade, and 3rd grade. Each child was asked to use color adjectives and size adjectives.

tives to describe an object that was given a nonce label. Some of the nonce words ended in *-a*, some ended in *-o*, and the rest ended in neither of these word markers. The bilingual children performed significantly better in marking target-like gender agreement for the masculine nonce nouns ending in *-o* and their agreeing adjectives than for the feminine nonce nouns ending in *-a*. These results are overall consistent with the results reported in Chapter 2 wherein adult heritage speakers are also more accurate with gender agreement on masculine nouns than feminine nouns, which has largely been taken to be in line with masculine as the default gender in Spanish.

Importantly, (Snyder et al. (2001)) found that there was no significant difference between the bilingual children in Kindergarten and 1st grade and their monolingual peers. However, the bilingual children in grades 2 and 3 performed worse on producing gender agreement than the bilingual children in Kindergarten and 1st grade. The authors argue that this evidence suggests language loss, specifically attrition of mastery of gender.

Anderson (1999) found similar evidence in a longitudinal study of the gender performance of English-Spanish bilingual siblings. The children's family had moved from Puerto Rico to the United States, at which point the sisters were primarily exposed to English in the community and at school. At the beginning of Anderson's study, the older sibling was 6;7 and the younger was 4;7. At this stage, the older sibling was producing gender agreement in the noun phrase without error, but after two years of living in the United States, her production of gender agreement decreased to 94.2% accuracy. The younger sibling also started with almost no gender errors, but by the last few sessions two years later had an error rate of up to 25%. Despite a sample size of only two children, Anderson took this to be evidence of language attrition in line with previous findings from (Snyder et al. (2001)).

Mueller Gathercole (2002) tested gender agreement in slightly older children: 2nd graders and 5th graders. Participants in this study were bilingual children of mixed socioeconomic status at a two-way bilingual and English immersion school in Miami; their gender agreement performance

was compared with that of age-matched monolingual children from Peru. Unlike other researchers, Mueller-Gathercole tested only non-canonical masculine and feminine nouns, under the assumption that gender assignment with canonical nouns ending in *-o* and *-a* would have already been acquired by the bilingual children at the age she was testing. (As Montrul et al. (2013) suggests, this assumption may need to be revisited given the reported gender performance of the children in Anderson (1999)'s longitudinal study.) Mueller-Gathercole reported that the bilingual children in her study performed below the level of the monolingual children in gender agreement, suggesting that their acquisition of gender was lagging – although more information on whether the children were sequential or simultaneous bilinguals may be necessary to determine whether the data could not also be interpreted as evidence of attrition rather than incomplete acquisition. Mueller-Gathercole did note that bilingual children who received greater amounts of Spanish-language input both at school and at home seemed to have an advantage over those bilingual students who had less Spanish-language exposure.

Montrul & Potowski (2007) conducted a study on a similar population at a Spanish-English two-way immersion school in Chicago. They considered gender agreement in bilingual children aged 6-8 and 9-11, as compared with age-matched L2s who attended the same school but spoke English at home, and as compared with age-matched monolingual children at a school in Mexico. Participants were asked to tell the story of Little Red Riding Hood, and their production of gender agreement on determiners and adjectives was recorded. In the second task, they were asked to complete frames of the type 'Veo...' ('I see') with the following schema: [determiner/numeral + animal + color] based on images they were shown.

In the oral narrative task, the older children performed at ceiling in production of gender agreement, while the younger children produced some errors. However, simultaneous and sequential bilinguals did not differ from each other. Taken together, this can be taken as evidence for no language loss with age, although the authors note that the amount of Spanish-language input in the

school environment put their bilingual participants at the highly proficient end of the heritage speaker spectrum.

Looking at the type of errors, the heritage speakers performed better on masculine gender agreement than on feminine gender agreement, again in line with previous work on children and adult heritage speakers of Spanish. There was a similar distribution of errors along gender lines in the second task for the heritage speakers, although there was no effect of age. The results showed some other notable asymmetries. First, the heritage speakers were more accurate on gender agreement on determiners than on adjectives. And second, the sequential bilinguals were much better in the second task than the simultaneous bilinguals. In fact, the simultaneous bilinguals did not quantitatively differ from age-matched L2 learners on the second task.

This body of research varies in methodology and target age group, but some trends are recognizable. At a very young age, bilingual children appear to not fall behind their monolingual peers in acquisition of gender (Müller (1987), Cantone (1999), Kuchenbrandt (2005), Eichler et al. (2013)). Nevertheless, older children tested show higher error rates in gender assignment and agreement, tentatively pointing to the susceptibility of gender to attrition. This susceptibility may at least in part be dependent on whether children are simultaneous or sequential bilinguals.

CODE-SWITCHING

Several investigations of code-switching among monolingual and bilingual children find striking asymmetries in the directionality and strategy of code-switching. The nominal domain appears to be the most frequent target of code-switching, especially for bilinguals speech (Poplack (1980), Myers-Scotton & Jake (2011), Cantone (2007), Eichler (2011)). Here, code-switching for bilinguals is almost strictly from the weaker heritage language to the stronger language, and functional elements are very rarely code-switched. Bilingual children frequently switch nouns from the majority language into Spanish, but that they typically use the Spanish articles (Licerias et al. (2008), Eichler et al. (2012)). In

addition, bilingual children have roughly equal preference for using the (default) masculine article before an English code-switched noun versus using the article that would be appropriate for the Spanish translational equivalent of the English word, while adult bilingual speakers show a strong preference for using the default masculine article with code-switched nouns.

3.1.3 PREVIOUS STUDIES ON GENDER IN HERITAGE SPANISH

Gender has been of particular interest to work on heritage languages for a few reasons. First, studies in L1 acquisition have reliably demonstrated that gender and gender agreement within the nominal phrase are acquired very early, and crosslinguistically are typically mastered by age 3 or 4 (see [Hernandez Pina \(1984\)](#), [Lleó \(1998\)](#), [López-Ornat \(1997\)](#), [Montrul \(2004\)](#), [Mariscal \(2009\)](#), [Pérez-Pereira \(1991\)](#) for Spanish; [Carroll \(1989\)](#), for French; [Rodina & Westergaard \(2012\)](#) for Russian; [Maratsos \(1988\)](#), for Polish, German, and Russian; [Gabor & Lukacs \(2012\)](#) for Hungarian; and [Comrie \(1999\)](#), for Isangu, a Bantu language; among many many others). Several years down the line, adult monolingual speakers make errors in gender agreement very rarely.

Nevertheless, heritage speakers of languages with grammatical gender consistently struggle with gender and gender agreement. Although instances have been reported in which some very high proficiency non-monolingual speakers of Spanish reach target-like accuracy rates in gender assignment and agreement ([Alarcón \(2011\)](#), [Kupisch et al. \(2013\)](#), [Montrul et al. \(2013\)](#)), for the most part even advanced heritage and L2 speakers fail to achieve monolingual-like mastery of grammatical gender, particularly when tested on noncanonical and exceptional nouns ([Franceschina \(2001\)](#), [Grüter et al. \(2012\)](#)). The results detailed below will demonstrate some consistent patterns in heritage performance with respect to gender: heritage speakers make more errors with feminine nouns than with masculine nouns, and more errors with noncanonical ending nouns than with canonical nouns ending in -o and -a. These errors will be evident regardless of the modality of the task, but some experimental methodologies (specifically the aural/oral ones) do give heritage speakers an advantage

in that they are able to perform closer to the level of monolingual speakers in such tasks (Alarcón (2011), Bowles (2011), Montrul et al. (2008)).

A majority of the work discussed below frames the discussion as a comparison between heritage speakers and L2 acquirers of Spanish, aiming to identify grammatical domains or task modalities that give one speaker group an advantage over the other. For the purposes of this project, what is most important is any divergence between the performance of monolingual speakers and of heritage speakers, and so the results of the papers below will mostly be discussed from this angle. For more detailed discussion of any implications for L2 acquisition, I refer the reader to the papers themselves.[§]

Most of the research on adult heritage speakers has focused on two primary research questions. First, to what extent does the reduced amount of input in acquisition to the heritage grammar (as compared to the monolingual grammar) cause performance differences between heritage speakers and baseline monolinguals? And second, to what extent does the fact that heritage speakers are still native speakers – by virtue of having acquired the language as children – give them an advantage as compared to those who acquire the language post-puberty?

Montrul et al. (2008) take as their starting point the notion that – although they do not quite perform at monolingual-level – heritage speakers should have an advantage over L2s because their acquisition of Spanish started in infancy. They tested heritage speakers on three different tasks, and compared with them monolinguals and adult L2 learners of Spanish. The first task involved written comprehension: participants were asked to read a sentence with noun drop, as in (1) below and select from among three items the one that was being discussed. In order to complete the task successfully, participants had to extract number and gender information from a determiner and/or any

[§]There is also a wealth of literature comparing gender performance of L2s directly to that of monolingual speakers. See Alarcón (2009), Keating (2009), Lew-Williams & Fernald (2010), Sagarra & Herschensohn (2011), Grüter et al. (2012) for a sampling of recent online measures of gender processing in L2s as compared to monolinguals.

nominal modifiers and use this to eliminate two of the given options. The heritage speakers struggled with this task (performing worse than both monolinguals and L2s), but did perform better when the task involved a masculine noun than when it involved a feminine noun.

- (ii) No quiero llevar las ___ de ese color.
NEG want.PRES.ISG bring DET.F.PL ___ of DET color
'I don't want to bring (those) in that color.'

In the second task, participants were presented with a written excerpt of a Spanish story or article, in which the articles had been removed. Participants had to choose the correct article to insert in each blank from two options they were given. This written gender recognition task required the extraction of number and gender information from the noun, in order to identify the correct determiner form. As with the previous written task, the heritage speakers performed significantly worse than baseline monolinguals, and also significantly worse than the L2 learners; accuracy was again higher on masculine than feminine nouns.

In the final task, participants saw pictures of items (labeled, but without a determiner, in order to control for dialectal variation in lexical items) and were asked to use them to complete the frame *Veo...* 'I see...' with a nominal phrase of the form [Determiner + Noun + Adjective]. In this oral production task, heritage speakers still did not perform at ceiling like the monolinguals did, but they did perform significantly better than the L2s learners. A much higher proportion of the heritage speakers had an accuracy rate above 80% than among the L2s. As in the previous tasks, accuracy rates were higher for masculine nouns than for feminine nouns, and participants were more accurate in conditions with canonical nouns ending in *-a* or *-o* than in conditions with non-canonical or exceptional nouns. An additional asymmetry was noted: heritage speakers (as well as L2s) were more accurate in producing gender agreement on determiners than on adjectives.

Montrul et al. (2008) determined that heritage speakers of Spanish clearly have non-standard performance in gender production, but that they do have a selective advantage based on task and

modality: they performed better in the oral production task than in the written tasks. The authors explain this modality asymmetry by referring back to the type of language acquisition experienced by heritage speakers. They interpret accuracy on the oral task as representative of implicit knowledge, while accuracy on written tasks is more indicative of explicit knowledge that requires some metalinguistic awareness. That heritage speakers struggle with written tasks and show an advantage in the oral task is then not surprising, given that implicit knowledge develops earlier for monolingual children (Bialystok (1994)) – at a time when heritage speakers are primarily exposed to Spanish, before a large increase in exposure to the dominant language – whereas explicit, metalinguistic knowledge does not develop until somewhere between age 3 and 5 and continues through formal schooling (Karmiloff-Smith (1979)). The authors conclude that heritage speakers do have competence in grammatical gender and agreement, but that the task modality and the type of linguistic knowledge that must be deployed affects heritage speaker performance due to the nature of their language acquisition experience.

Alarcón (2011) replicated the written gender recognition task and the oral picture description task from Montrul et al. (2008) but with highly proficient adult heritage speakers of Spanish. Crucially, these speakers were all sequential bilinguals. In the written task, Alarcón (2011) found that the high proficiency heritage speakers performed much better than they had in Montrul et al. (2008) – still not target-like, but on par with the L2 learners, indicating some level of explicit knowledge of Spanish. They did also replicate the asymmetry between performance on masculine versus feminine nouns, on articles versus adjectives, and on canonical versus non-canonical ending nouns. Alarcón (2011) modified Montrul et al. (2008)'s oral production task to eliminate all exception nouns and animate nouns. The results replicated the findings in Montrul et al. (2008) that heritage speakers performed better than L2s but still not at the level of the monolinguals. They also again replicated the same distribution of error types as in the written task and as in the tasks in Montrul et al. (2008). The asymmetries between masculine and feminine in particular support the notion that the mas-

culine is the default gender feature in Spanish. Bringing together evidence from the written task and the oral task, Alarcón (2011) concludes that high proficiency heritage speakers do have both explicit and implicit knowledge of gender in Spanish, which could suggest complete acquisition of this grammatical feature. Under this analysis, the diverging accuracy rates between heritage speakers and baseline monolinguals would be attributed to performance issues rather than differences in competence or representational differences.

Drawing on previous work that seems to indicate heritage speakers do overall have some advantage in gender acquisition over L2s, particularly when it comes to non-canonical ending nouns, Montrul et al. (2013) tried to separate type of acquisition from age of acquisition as the crucial distinguishing factor. They identified diminutives as a property of child language acquisition that is almost absent from adult language use: Marrero et al. (2007) found that the frequency of diminutives in child language and Child Directed Speech in peninsular Spanish is as much as 13 times higher than in adult speech.

Incidentally, diminutives are transparently marked for gender: they must always end in *-o* or *-a*, so they facilitate the acquisition of gender agreement for nouns with exception of noncanonical endings (Kempe & Brooks (2001), Savickienė & Dressler (2007), Seva et al. (2007)). Consider for instance two nouns with the noncanonical ending *-e* – *el coche* ‘the car, masc.’ and *la nube* ‘the cloud, fem.’ In the diminutive, the gender of both of these nouns becomes transparently marked through canonical endings: *el cochecito* and *la nubecita*, respectively. Montrul et al. (2013) predicted that heritage speakers should have a better familiarity with diminutive forms and be more accurate at producing gender agreement with diminutives, because these were a part of the language input to their grammar in first language acquisition through Child Directed Speech – as is the case for baseline monolinguals, but what distinguishes them from L2 acquisition.

Participants in the Montrul et al. (2013) study were asked to provide the diminutive forms of a series of nouns in a preliminary task, and then were given an oral elicitation task. Consistent with

results from previous oral elicitation tasks, Montrul et al. (2013) found that heritage speakers were not quite target-like in their oral production – they performed just below at-ceiling – but certainly produced gender agreement better than the L2s did, particularly for the non-canonical nouns. They were also more accurate with masculine nouns than with feminine nouns. Contra expectations, Montrul et al. did not find evidence that diminutives boosted heritage speaker performance in gender agreement, although they did find that heritage speakers were more familiar with diminutive forms in the preliminary task than the L2s were.

Montrul et al. (2014) continued to investigate how the particular type of language acquisition experience of heritage speakers impacts their performance in gender agreement as adults. Again, the starting observation is that heritage speakers have access to the same cognitive, linguistic, and processing mechanisms specific to L1 acquisition that are assumed to be absent from L2 acquisition, but their acquisition trajectory diverges from that of baseline monolinguals around school age, when monolingual children begin to receive explicit instruction in the language, often in written form. They asked whether the early language experience that heritage speakers share with monolingual children gives them any advantage in performance in gender morphosyntax, which is acquired early in language development, as discussed above. Specifically, they were hoping to identify tasks in which heritage speakers' behavioral patterns were closer to that of the monolingual baseline.

Monolingual, heritage, and L2 participants in the study performed three tasks: an aural/oral Word Repetition Task (WRT); an aural Gender Monitoring Task (GMT); and an aural Grammaticality Judgment Task (GJT). Stimuli included both canonical and non-canonical Spanish nouns; as the authors argue, target-like knowledge of gender requires knowledge of the gender of non-canonical nouns as well. All three tasks were presented aurally, as this had previously been determined to boost heritage speaker performance, but they varied in the explicitness of the task. The GMT and GJT required that participants attend to gender directly (through making a judgment on the gender of a single word in the GMT) or indirectly (through making a judgment regarding

the gender agreement in a phrase in the GJT), whereas the WRT tested more implicit and automatic knowledge of gender. Montrul et al. predicted that if heritage speaker performance is only a matter of their type of language input experience (oral/auditory, rather than written), then they should perform equally well on all three tasks, and be more target-like than L2s. If, however, the type of linguistic knowledge that needs to be deployed (implicit rather than explicit, as suggested in Bowles (2011), Rebuschat & Williams (2012)) also determines their performance, then only the WRT should yield closer to target-like performance for the heritage speakers, while the remaining two tasks should result in more divergence between baseline and heritage performance.

Consistent with the hypothesis that the degree of explicitness of linguistic knowledge at least partly determines heritage speaker performance, the heritage speakers performed worse on the explicit GMT and GJT, but target-like in the WRT (baseline monolingual speakers performed at ceiling on all three tasks). The heritage speakers also had more difficulty with non-canonical nouns, but Montrul et al. suggest that this may be due to the lower frequency of non-canonical nouns in the lexicon (Gollan et al. (2011)). Crucially, although their performance was non-target-like on non-canonical nouns, heritage speakers still demonstrated a sensitivity to grammaticality with these lexical items. Montrul et al. (2014) suggest that the advantage heritage speakers have in tasks that deploy implicit knowledge of gender stems from properties of first language acquisition, rather than just age of acquisition: early speech segmentation in child language learning facilitates the development of a close association between the determiner and the noun, and aids gender learning.

Scontras et al. (2018) considered whether performance differences between heritage speakers and monolingual speakers may be symptomatic of differences between monolingual and heritage speakers in the mental representation of the grammar of Spanish. The authors replicated the study in Fuchs et al. (2015) to determine what the structure of number and gender features on the nominal spine is for heritage speakers, bearing on the discussion in Section 1.4. Fuchs et al. (2015) had concluded that number and gender are projected and valued independently in monolingual Spanish

DPs. Their experimental design involved manipulation of ungrammatical sentences with agreement attraction errors. Monolingual speakers should show a cumulative effect of errors in number and gender agreement: sentences with errors in both features on a single predicative adjective were more degraded than sentences with an error in just one feature on the predicative adjectives. The authors took this to be evidence that agreement in number and gender proceeds independently of the other, suggesting the two features are split in the syntactic representation.

In the follow-up, [Scontras et al. \(2018\)](#) found that heritage speakers found errors in both number and gender to be equally degraded as errors in just one of the features. Therefore, the authors argue that number and gender agreement in the heritage grammar is a single process, and that number and gender are bundled together on a single projection in the syntax. They therefore suggest that the divergent performance on gender agreement between monolingual and heritage speakers is at least in part the result of divergent knowledge of the grammar.

What has been consistently observed in heritage linguistics is a difference in performance in gender and gender agreement between monolingual baseline speakers of Spanish and heritage speakers of Spanish. The research discussed in this section has been largely preoccupied with determining what might affect the performance of heritage speakers – by either boosting it or weakening it. These questions in turn provide some insight into the deeper question of why are those differences in performances there in the first place. The variety of approaches is reflective of a variety of factors that might be at play in shaping heritage speakers' performance in gender production and comprehension, including but not limited to task modality, explicitness of the type of linguistic knowledge being targeted, representational differences in the underlying grammar of heritage speakers, and the nature of language input and language experience that was available to the heritage speaker during first language acquisition.

3.1.4 PREVIOUS STUDIES ON PREDICTIVE USE OF GENDER IN SPANISH

A relatively recent line of work investigates whether grammatical gender aides in comprehension in online processing tasks. Research in this domain has made use of eye-tracking in the Visual World Paradigm (VWP) to determine the influence of gender information on spoken word recognition. The idea behind these studies is relatively simple. When presented with images of two objects of the same grammatical gender (let's say both are feminine: *la manzana* 'apple' and *la camisa* 'table'), upon hearing the sentence *Dónde está la manzana?* 'Where is the apple?', participants must wait until the onset of the lexical item (the phoneme [m]) to be able to identify the target item and fixate on the image of the apple. By contrast, when presented with one masculine image (let's say, *el libro* 'book') and one feminine image (let's use *la manzana* 'apple' again), participants are provided with disambiguating information before the onset of the lexical item in the question *Dónde está la manzana?*, in the form of the article *la*. In other words, in this "mismatch" condition, the gender information contained in the article *la* is sufficient to minimize the set of possible candidates earlier in the sentence and anticipate the target item.

This effect has been found to be quite robust for monolingual speakers of several European languages (Paris et al. (2006) for German; Lew-Williams & Fernald (2007) for Spanish; Loerts et al. (2013) for Dutch; and Hopp (2013) for German). For Spanish-speaking adults, Lew-Williams & Fernald (2007) found that adults on average shifted their gaze to the target item faster in mismatch conditions (approx. 600 ms after the onset of the article) than on match conditions, in which the gender of both images on the screen was the same (approx. 700 ms after the onset of the article), as shown in Figure 3.1. While the authors do not say so explicitly, their example stimuli suggest the lexical items used in the study had canonical *-o/-a* word markers. The authors performed the same study on children (age 34-42 months) and found a similar effect: children oriented to the target item faster on mismatch conditions (approx. 850 ms after article onset) than on match conditions (ap-

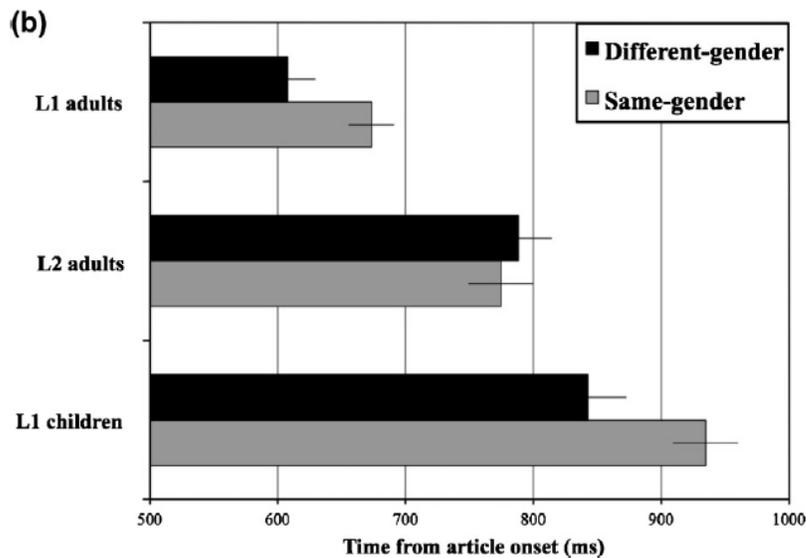


Figure 3.1: Mean response time of Spanish-speaking L1 adults, L2 adults, as well as L1 children in Lew-Williams and Fernald (2007, 2010).

prox. 950 ms after article onset), indicating that even at age 3, monolingual[¶] Spanish children are able to use gender information earlier in the sentence to facilitate spoken word recognition. Note however that the children were slower across the board than the monolingual adults. For both adult and child participants, the effect was present both when the target noun was masculine and when it was feminine, and the authors do not report any significant difference between genders.

In a follow-up study, [Lew-Williams & Fernald \(2010\)](#) compared L1 speakers of Spanish to L2 speakers of Spanish in the same experimental setup. In contrast to the population from the 2007 study, which included primarily participants of low socioeconomic status and on average nine years of education, the participants in this study were all university students. The L2 group were all L1 speakers of English. Overall, the L2 speakers performed slower than the monolingual adults and faster than the monolingual children (monolingual results replicated the findings in [Lew-Williams](#)

[¶]Child participants were the children of recent Latino immigrants from Mexico to the United States. Based on a questionnaire assessing language input, the authors argue the input to these children's grammar was about 85% Spanish. One might argue that some of them could be considered bilingual children.

& Fernald (2007); child results drawn from Lew-Williams and Fernald 2007), but their speed was about equal for match and mismatch conditions (Figure 3.1). On the basis of these results, the authors argued that L2 speakers of Spanish are not able to use gender information to facilitate word recognition, setting them apart from L1 children and adults, who have an advantage in processing gender information on determiners.

Noting that the L2 participants in Lew-Williams and Fernald (2010) might be considered intermediate-proficiency speakers of Spanish, Grüter et al. (2012) set out to test high proficiency L2 speakers of Spanish in the same experimental paradigm. Their results confirmed those of Lew-Williams & Fernald (2007, 2010), in that L1 adults were again able to use gender to facilitate noun retrieval, whereas L2 speakers – in this case, highly proficient L2 speakers – were not.

To explain the differences between adult and child monolingual speakers of Spanish on the one hand, and L1-English, L2-Spanish speakers on the other hand, Grüter et al. (2012) considered differences in the language-learning experience of these two groups. L1 speakers of Spanish learn the language from naturalistic input from the speech stream, in which article + noun sequences are incredibly frequent. In fact, studies on child language acquisition suggest children first treat article + noun combinations as a single unit or chunk, only subsequently analyzing them as composed of two separate morphemes, possibly with an intermediate stage in which the article is analyzed as a prefix on the noun (Carroll (1939), MacWhinney (1978), Pinker (1984), Abu-Akel et al. (2004), Basano et al. (2008), Mariscal (2009), Pine & Lieven (1997), Tomasello (2000)). Building on a similar discussion in Lew-Williams & Fernald (2010), Grüter et al. suggest that this learning experience is conducive to forming a “tight link” between the article and the noun, that enables L1 speakers from a very early age to use information on articles to facilitate spoken word recognition. By contrast, L2 learners of Spanish do not have nearly as much naturalistic input to their grammar. Most of their experience with Spanish involves written forms of the language, and these include metalinguistic information, such as spaces between article and noun, that are less likely to foster a tight association

between articles and nouns. This, the authors argue, may explain why L2 speakers are unable to use grammatical gender information on determiners to facilitate word recognition.

Dussias et al. (2013) also replicated the study in Lew-Williams & Fernald (2007, 2010) on L1 adults and highly proficient L2 speakers (they had three groups: highly proficiency L1-English, L2-Spanish adults; lower proficiency L1-English, L2-Spanish adults; and L1-Italian, L2-Spanish adults). In their design, most feminine target items ended in the canonical word marker *-a* (93%), while masculine target items ended in *-o* (69%), *-e* (12%), and some consonants. They also departed from the earlier studies in using more complex carrier sentences, half of which had the target item embedded in the middle of the sentence (12-a), while in the other half the target item appeared at the end of the sentence (12-b).

- (12) a. El estudiante estaba dibujando el reloj que vio ayer.
the.M student be.IMP.3SG draw.PROG the.M clock.M that see.PRET.3SG yesterday
'The student was drawing the clock that he saw yesterday.'
- b. El niño miraba a su hermano mientras fotografiaba el reloj.
the.M boy.M look.IMP.3SG at his brother while photograph.IMP.3SG the.M clock.M
'The boy watched his brother taking a picture of the clock.'

In contrast to the findings by Grüter et al. (2012), Dussias et al. (2013) found that highly proficient L1-English, L2-Spanish adults did fixate on target items faster on mismatch conditions than on match conditions, suggesting second-language learners of Spanish may be able to use gender information to facilitate spoken word recognition. For lower proficiency L1-English, L2-Spanish adults, they did not find this effect. In fact, they found the opposite effect when the target item was masculine: participants fixated on masculine target items slower in the mismatch conditions than in the match conditions.

In light of earlier findings, Dussias et al. (2013) certainly complicate the picture a bit with respect to whether L2 speakers of Spanish do or do not pattern with (effective) monolinguals. It is worth

noting here again that Dussias et al. employed a different methodology than their predecessors, one that entailed a higher cognitive load and that makes a direct comparison of their findings slightly difficult. The presence of the opposite effect for masculine target items for L2 speakers in their study is also very unexpected, and deserves a more detailed explanation, although they do replicate the findings for monolingual speakers of Spanish from earlier studies (Lew-Williams & Fernald (2007, 2010), Grüter et al. (2012)).

Taking their results at face value though, Dussias et al. (2013) offer a competing hypothesis to the “tight link” explanation offered by Grüter et al. (2012). They suggest that the explanation is more likely to be grounded in language dominance. More dominant/proficient speakers (monolinguals or second-language learners) have more resources available to attend to various linguistic elements during online processing, and are therefore more likely to be able to identify determiners marked for gender and use them to facilitate recognition of subsequent nouns. In contrast, less proficient second-language learners have limited resources during online processing and are likely to prioritize other linguistic elements or be unable to access gender information in real time.

Heritage speakers offer the opportunity to test between the competing explanations offered by Grüter et al. (2012) and Dussias et al. (2013). By definition, heritage speakers are those who learned the language at home, giving them the same language learning experience as monolingual children and adults. This language experience should be conducive to forming the kind of “tight link” argued for by Grüter et al. and further supported by evidence from Montrul et al. (2014), discussed previously, but this has not been tested using such a sensitive measure as eye-tracking. On the other hand, heritage speakers are no longer dominant in their L1. Despite their shared language learning experience with (effective) monolingual children and adults, adult heritage speakers have in common with L2 speakers of Spanish the fact that Spanish is not their dominant language, and they share certain patterns in performance of gender production. Their lower proficiency (as compared to monolingual adults) might cause them to be unable to use gender to facilitate lexical retrieval, if the proposal

offered by [Dussias et al. \(2013\)](#) is on the right track.

The present study therefore aims to answer the question “Can heritage speakers use gender information on determiners to facilitate spoken word recognition?” The experimental design replicates that of [Lew-Williams & Fernald \(2007, 2010\)](#) and [Grüter et al. \(2012\)](#), thus ensuring maximum room for comparison with previous findings. The hypotheses are as follows:

(13) Hypothesis 1: If the ability to use gender to facilitate word recognition is associated with the nature of the learning experience – the so-called “tight link” – then heritage speakers should pattern with L1 adults and children in being able to use gender information on determiners to anticipate the subsequent nouns.

Hypothesis 2: If this ability is instead tied to language proficiency, then heritage speakers should pattern with L2 learners who, in the [Lew-Williams & Fernald \(2010\)](#) and [Grüter et al. \(2012\)](#) experimental setup, were unable to use gender on determiners to anticipate subsequent nouns.

In the next sections, I will detail the experimental design and results of a study I ran to test these two hypotheses, and suggest further steps that may allow us to further probe the nature of the process we are observing in these studies.

3.2 MATERIALS AND DESIGN

This section will detail the design of the eye-tracking study, the goal of which is to replicate the methodology in [Lew-Williams & Fernald \(2007, 2010\)](#) and [Grüter et al. \(2012\)](#). The selection of images and creation of visual stimuli had to control both for standard concerns of experimental design as well as the fact that Spanish is spoken in many different countries around the world, resulting in dialectal differences not just in syntax and phonology but also lexicon. How this was controlled

for will be described in 3.2.1 and 3.2.2. Considerations relating to the design of auditory stimuli to ensure time for anticipatory looks are detailed in 3.2.3.

3.2.1 IMAGE SELECTION AND NORMING

A set of 175 images previously used in eye-tracking studies by Holger Hopp and his colleagues was obtained for the purposes of this study. In order to determine which of these images would be useable in a Spanish-language study, I ran a norming study to control for recognizability, consistency in choice of lexical item, and dialectal consistency in gender assignment.

The images were arbitrarily divided into four groups. Each group was posted as an independent Human Intelligence Task (HIT) on Amazon Mechanical Turk. In order to discourage participation of non-native speakers of Spanish, the title and instructions for the HIT were in Spanish, and the IP addresses of contributors were restricted to those within the United States and Latin-American Spanish-speaking countries. Participants were also asked to provide their age and country of birth and to self-report their proficiency in speaking, reading, listening, and writing in Spanish.

For each image in the HITs, participants were asked to write the Spanish name of the object, using the article *el* or *la*. 30 contributors completed this task for each of the sets of images, for a total of 120 contributors, who were all compensated for their time. Answers were removed for those participants who reported a country of birth that is not a Spanish-speaking country or the US, and for participants who reported low measures of proficiency. This left a total of 93 contributors ($n=22$ for the first set of images, $n=23$ for the second set, $n=22$ for the third set, and $n=26$ for the fourth set).

Each image was then checked for three key properties. The first was recognizability. I eliminated any image for which participants did not consistently identify what it was meant to represent. For example, the image representing a bookshelf had to be eliminated, because a not-insignificant fraction of the participants labeled the image as *los libros* ‘books’, thinking the image represented books rather than the bookshelf. Second, images had to show lexical consistency. Thus, any image for

which there was significant variation in lexical assignment had to be eliminated. For instance, the image of the ball was labeled by half of the contributors as *la pelota* ‘ball’ but as *el balon* ‘ball’ by the other half. This image was not useable. Similarly, the image of the pot was labeled as *la olla* ‘pot’ by some of the participants, while others labeled it as *la maceta* ‘pot’. Although these words have the same gender, experimental items need to have little-to-no dialectal variation in both gender and lexical item, so the pot was also not usable. Finally, I controlled for consistency in gender assignment. There were some images for which the different lexical items assigned were nearly identical but differed in gender assignment. For instance, the image of the purse was alternately labeled as *la bolsa* (feminine) and *el bolso* (masculine). Because it is crucial that experimental items have a consistent gender, the image of the purse was not usable.

Three other considerations for experimental design eliminated some of the images based on factors not related to dialectal variation. Those for which the Spanish label is a cognate of the English term were eliminated, since it might be debatable which language’s lexicon a speaker would be accessing upon hearing the lexical item. Second, only lexical items whose first phoneme is a consonant and not [l] were included. The phonology of Spanish is such that co-articulation between *el/la* and such lexical items would make it impossible to identify a clear word boundary. Finally, in order to ensure enough time for eye-movements, words with one syllable were considered potentially too short and were therefore eliminated.

After filtering out images based on these requirements, I selected an equal number of images for each gender: 21 feminine, 21 masculine. These images, along with their frequencies in child-directed speech, are listed in Appendix A.1. Within each gender, they were equally distributed according to color: 7 blue images, 7 red images, 7 green images. Although color does not play a role in the auditory stimuli in this study in Chapter 4, it does so in the Polish study, so it was important to ensure consistency in its use here as well.

3.2.2 VISUAL STIMULI

Each visual stimulus consisted of a white background, with one image on the left and one on the right, equidistant from the center. A fixation cross was placed in the center of the display.[‡] 42 of the visual stimuli were control “match” conditions, in which the gender of the target item was the same as that of the distractor item (Figure 3.2). 42 of the visual stimuli were test “mismatch” conditions, in which the gender of the target item was different than that of the distractor (Figure 3.3). An additional 18 “match” conditions were added in anticipation of the German iteration of the study, conducted at a separate time and location, with additional considerations regarding the gender of the German equivalent of the two images with matching Spanish genders. The stimuli were balanced such that target items were equally balanced between masculine and feminine gender. Each of the 42 images appeared two or three times as the target item and two or three times as the distractor. Whether the target item was on the left or the right of the screen was determined by a randomization function.

In combining images into pairs, there were two important considerations. The first takes into account evidence from eye-tracking studies that some colors may attract eye-movements faster than others (see discussion in [Huettig et al. \(2011\)](#), [Huettig & Altmann \(2011\)](#), [Mani et al. \(2013\)](#)). Therefore, both images in a visual stimulus had to be of the same color.^{**} This also anticipates the experimental design of the Polish study and ensures consistency in the designs of the two studies. The second consideration is more linguistic in nature. Images had to be paired such that the first phoneme of the two words was different. The goal is for the onset of the lexical item to clearly disambiguate

[‡]This is unlike the studies in [Lew-Williams & Fernald \(2007, 2010\)](#). Because their study involved children, who are unlikely to be able to follow instructions regarding orienting on a fixation cross, they chose to omit this in their visual stimuli. Their participants could engage freely with the VWP, which forced them to eliminate about half of the data, in which participants had already been oriented on the target item at the moment in which the article and target item began to unfold in the auditory stimulus.

^{**}Special thank you to Hania Bigoszevska for working with the visual stimuli to change their color and clarity as needed.



Figure 3.2: Spanish Match Condition: The target item *el casco* 'helmet' and the distractor *el martillo* 'hammer' are both masculine.



Figure 3.3: Spanish Mismatch Condition: The target item *el casco* 'helmet' is masculine, while the distractor *la vela* 'candle' is feminine.

between the two items in each stimulus. So for instance, *la pluma* ‘feather’ and *el puño* ‘fist’ could not be paired.

3.2.3 AUDITORY STIMULI

Each target item was placed in a sentence with the structure in (14), similar in brevity and simplicity to the auditory stimuli in [Lew-Williams & Fernald \(2007, 2010\)](#) and [Grüter et al. \(2012\)](#), and different from those in [Dussias et al. \(2013\)](#). All sentences were recorded by a male speaker of Colombian Spanish recently immigrated to the United States. The speaker spoke at a slow pace and read each sentence with the intonation of a question.

(14) ¿Dónde está ARTICLE NOUN?

In order to ensure that the onset of the article and the lexical item was exactly the same in all auditory stimuli, the recorded sentences were spliced at the boundaries of each key phrase: the frame, the article, and the noun. A single token of the frame was chosen to be the frame for all auditory stimuli (length = 1150 ms); for the article, a single token of “el” (length = 280 ms) and a single token of “la” (length = 280 ms) of the same length were chosen. The average length of lexical items was 585 ms. This splicing also obscured any coarticulation between articles and nouns, which might have otherwise provided a cue to the lexical item in addition to the gender information on the article.

The auditory stimulus for each target item was formed by concatenating the frame, the appropriate article *el/la*, 200 ms of silence, and the target item itself. Total, there was 480 ms between the onset of the article (the provision of gender information) and the onset of the lexical item. Each auditory stimulus was preceded by 200 ms of silence and a 600 ms high-pitched beep and a further 200 ms of silence. The purpose of this time before the onset of the auditory stimulus was to give participants time to familiarize themselves with the objects on the screen, which makes them more likely to be able to make anticipatory eye movements. Participants were instructed that they could look



Figure 3.4: Spanish auditory stimuli were spliced together to ensure that the onset of the gender information and of the lexical item was at the same time in each stimulus.

at the items on the screen when they first appeared, but that they had to direct their gaze to the fixation cross once they heard the beep. A schematization of the duration of each part of the auditory stimulus is in Figure 3.4.

There are two useful outcomes of this careful design of the auditory stimuli. The first is that the onset of the article and the onset of the lexical item occurs at the same time in each trial. This allows for ease of comparison in the statistical analysis in Section 3.7. Second, because participants are fixated on the cross at the onset of the auditory stimulus, it is highly unlikely that they are already looking at the target item or distractor when they hear the onset of the article or lexical item (if they were, the data from such a trial would be meaningless, since it would not be indicative of anticipatory looks but rather luck), unless a participant did not follow directions.

3.3 MODEL INTERPRETATIONS

In keeping with the analyses conducted in previous studies (Lew-Williams & Fernald (2007, 2010), Grüter et al. (2012), Dussias et al. (2013)), the dependent variable of interest is the time of first fixation on the target item. This measure indicates the moment in which participants are able to shift from the fixation cross to orient on the target item, having minimized the set of candidate lexical items. If participants are able to use grammatical gender information on articles to anticipate subsequent lexical items, then they should orient to the target item faster in mismatch conditions, in

which the gender on the article carries sufficient information to disambiguate between the two candidates.

(15) Model interpretations:

1: If heritage speakers can use gender information to facilitate lexical retrieval, then the time of first fixation on target items will be faster in mismatch conditions than in match conditions.

2: If heritage speakers cannot use gender information to facilitate lexical retrieval, then the time of first fixation on target items will be about equal in match and mismatch conditions.

3.4 PARTICIPANTS

Forty-five Spanish speakers aged 18-65 from the Cambridge and Boston area were recruited for the study. Participants were compensated for their time. They were asked to fill out a demographics and language use questionnaire, which was an abbreviated version of the Language Experience and Proficiency Questionnaire (LEAP-Q).^{††} The LEAP-Q was available to be filled out either in Spanish or in English, and participants were encouraged to fill out the form in whichever language they felt most comfortable with.

The LEAP-Q contains questions pertaining to language upbringing (years spent in a Spanish-speaking family, country, and school), current language use, and self-reported language proficiency in reading, understanding, and speaking (on a scale of 1-10). The LEAP-Q also asks participants to self-report how much they perceive their own accent in Spanish to be “non-native” (on a scale of 0-10) and how frequently others identify them as “non-native” speakers of Spanish (on a scale of 0-10).

^{††}The original version of the LEAP-Q is available – in many languages – at <https://bilingualism.northwestern.edu/leapq/>. The abbreviated version used for the purposes of this experiment is attached in Appendix B.1 (English) and B.3 (Spanish).

Participants were identified as belonging to the heritage group or control group based on the number of years they spent in a Spanish-speaking country. Those who spent 8 years or less in a Spanish-speaking country were considered to be heritage speakers, since they would have immigrated to the United States and started learning English before the onset of puberty. Participants who had spent 18 years or more in a Spanish-speaking country were considered to be the control group. For cleanliness of data, those who spent 9-17 years in a Spanish-speaking country were excluded from the analysis. Given that the onset of bilingualism occurs during or after puberty for these participants, and most likely after the acquisition of most if not all parts of the grammar of the L1, they are considered (early) L2 learners/speakers (sometimes referred to as the “1.5-generation”), whose linguistic ability falls outside of the scope of interest in this dissertation (Unsworth (2005)).

A few other participants were excluded for miscellaneous reasons. Two participants were over the age of 60, and were excluded for general concerns regarding language attrition in older populations. Two participants were excluded because they reported that the first language they learned was neither Spanish or English (these were L1 speakers of Catalan and Japanese). One participant was excluded from analysis because this person fell asleep during the study. Another participant reported an ear infection, which may have impacted their sensitivity to the auditory stimuli. One further participant was excluded because this speaker’s answers to the demographics questions were contradictory, making it impossible to classify them as heritage or control.

Given all the exclusions, the final sample size was $n=21$ for the heritage speakers, and $n=10$ for the controls. Figure 3.5 shows the age distribution of the participants, and Figure 3.6 shows a boxplot of the length of time that participants in each of the two groups spent living in a Spanish-speaking country.

Table 3.2 provides an overview of the demographic data for the heritage and control group. These data were gathered from responses to the LEAP-Q. As is typical of heritage speakers, the participants identified as heritage speakers in this study spent on average very little time in a Spanish-

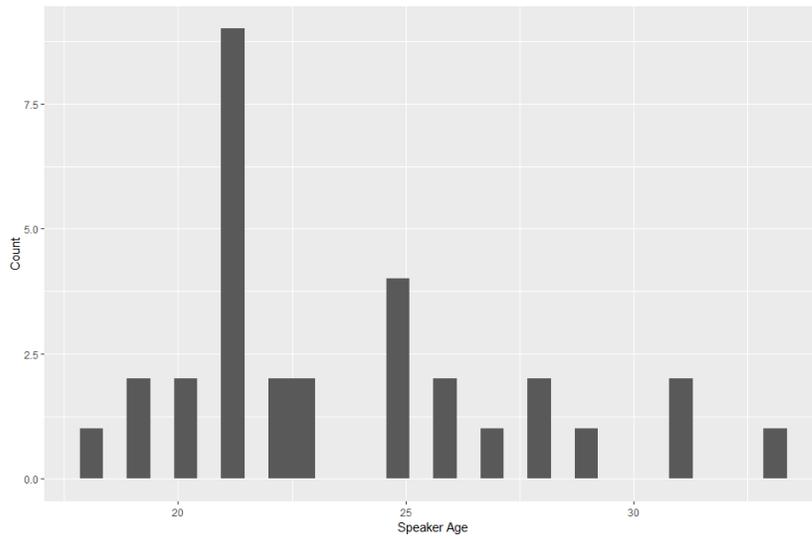


Figure 3.5: A histogram of the ages of Spanish-speaking participants in the study whose data was included in the analysis.

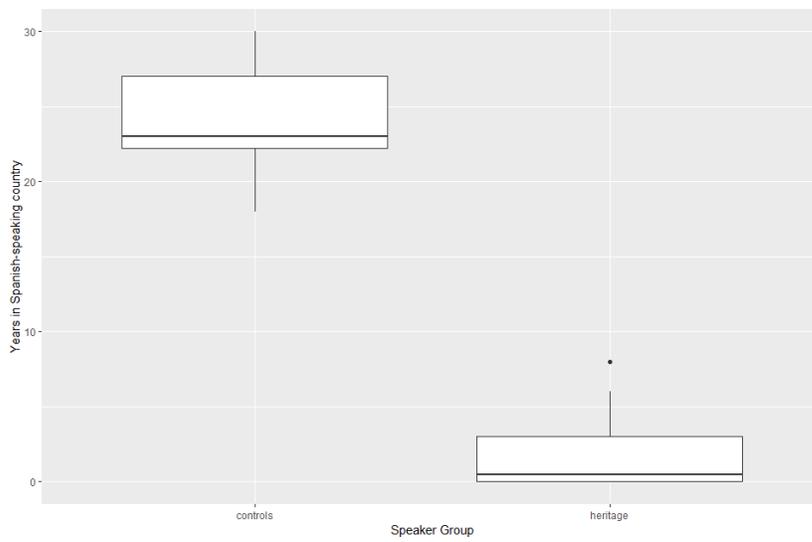


Figure 3.6: The number of years that participants in the heritage group and the control group reported living in a Spanish-speaking country.

	<i>n</i>	Age	Years in Spanish-speaking environment			
			Country	Family	Work/School	
Controls	10	26.4 (4.0)	23.2 (4.5)	20.9 (8.2)	20.2 (8.6)	
Heritage	21	22.3 (3.2)	1.9 (2.8)	16.0 (9.6)	3.0 (4.8)	
	Education level					
	HS	Some Coll.	College	Some Grad	Masters	PhD
Controls	1	1	2	1	4	1
Heritage	0	10	6	3	1	1

Table 3.2: Demographic data for Spanish-speaking participants, collected from the LEAP-Q. Education level indicates the highest completed degree.

speaking work and school environment, but did spent most of their lives in a Spanish-speaking family.

3.5 PROCEDURE

Participants were tested individually in a lab, where they first filled out the LEAP-Q. They then completed a lexical production task (more on this in Section 3.6), which was displayed on a laptop, and the audio was recorded in Praat. They then completed the eye-tracking comprehension task. Participants sat facing a 53.5-cm screen, approximately 75cm away, with their heads placed in an apparatus that ensured minimal head movement during the task. Participants received instructions in written and oral form. They completed four practice trials before the calibration of the SMI Eye-link 1000; the goal of calibration was set at achieving visual acuity below 0.5 degrees. Gaze position was recorded at 2000Hz. Participants were allowed to take a break midway through the study, after 51 trials. The length of the break was determined by the participants, and calibration was repeated before the start of the second half of the trials. In total, the study took approximately 30-45 minutes depending on ease of calibration of the eye-tracker and on how long participants chose to take break. Participants were compensated for their time.

3.6 PROFICIENCY

The lexical production pre-task served two purposes in this study. One of these two purposes was to clean the data to ensure more representative results, which I will address in 3.7. The other purpose was to have a measure of the lexical proficiency of the participants. Many studies on L2 and heritage speakers of Spanish use the DELE exam (Diplomas de Español como Lengua Extranjera) to assess proficiency. However, the DELE is a written test of linguistic ability, which puts heritage speakers at a disadvantage (Montrul et al. (2008)) and it is time-consuming. On the other hand, the lexical production task involved oral picture-naming, which heritage speakers have been shown to perform better on (Montrul et al. (2008)). This lexical production task was already necessary to ensure representative results, and findings that show lexical proficiency can serve as a reliable proxy measure for syntactic proficiency (Polinsky (1997, 2006), Godson (2003)) allow us to take advantage of this task for assessing speaker proficiency as well.

In this particular lexical production task, participants viewed a powerpoint presentation in which each slide contained one of the 42 images that they would see in the subsequent eye-tracking task. They were asked to identify each item orally in Spanish, using an article and a noun. The task was self-paced, and participants were told they could think as long as they needed before giving a response. They were also told that if they could not recall the name of a particular item, they could skip it. Responses were recorded in Praat.

One important choice regarding the oral recognition task was its timing relative to the rest of the experimental procedure. It is generally accepted that any kind of pressure put on vulnerable populations in an experimental setting might impact their performance on the task. Heritage speakers are one such population, and a task in which they must label (or fail to label) images in their heritage language could be such a task (failing to identify an image could make a participant hyper-aware of their divergent knowledge of the language). This would suggest that it might be more strategic to

have participants encounter the lexical production task after the main eye-tracking task. However, the ultimate determining factor in having the lexical production task come before the eye-tracking task was the issue of “re-learning”. Participants might re-learn the vocabulary items during the eye-tracking task, which would lead to inflated – not representative – results on the lexical production task if it were to occur after the eye-tracking portion. In weighing the options, this was deemed to be a more significant problem than any effects of frustration resulting from the pre-eye-tracking lexical production task. To mitigate any such effects, experimenters (either me or my research assistants) were instructed to put participants at ease by signaling that their ability to name all of the images would in no way affect their eligibility to participate or the results of the study, and that it was very common for participants in the study to forget some fraction of the words.

To obtain a measure of lexical proficiency, participants’ responses were identified as correct if they were an appropriate label for the item in any Spanish dialect,^{‡‡} preceded by the correct article (*el/la*). For each participant, I recorded the percent of vocabulary items that they labeled correctly. The results, split by heritage versus control group, are presented in Figure 3.7.

As is evident from Figure 3.7, the control group performed at ceiling. With a few exceptions, participants were able to correctly identify all of the images. Meanwhile, the heritage speakers overall knew less of the vocabulary items (average 34 out of 42), and they showed much more variability (range 19-42 out of 42): less proficient speakers knew as little as about half of the lexical items, while more proficient speakers fell in the same range as the controls.

Of the total 168 words that the heritage speakers were unable to label correctly, only 8 were coded as incorrect due to a non-target determiner and/or word marker that might suggest incorrect gender assignment (ex. **el palo*, intended: *la pala* ‘shovel’; or **el velo*, intended: *la vela* ‘candle’). The remaining missed items were coded as incorrect because participants did not provide an answer or pro-

^{‡‡}In contrast with what will be seen in 3.7, this label did not have to match the one intended by the experimental design.

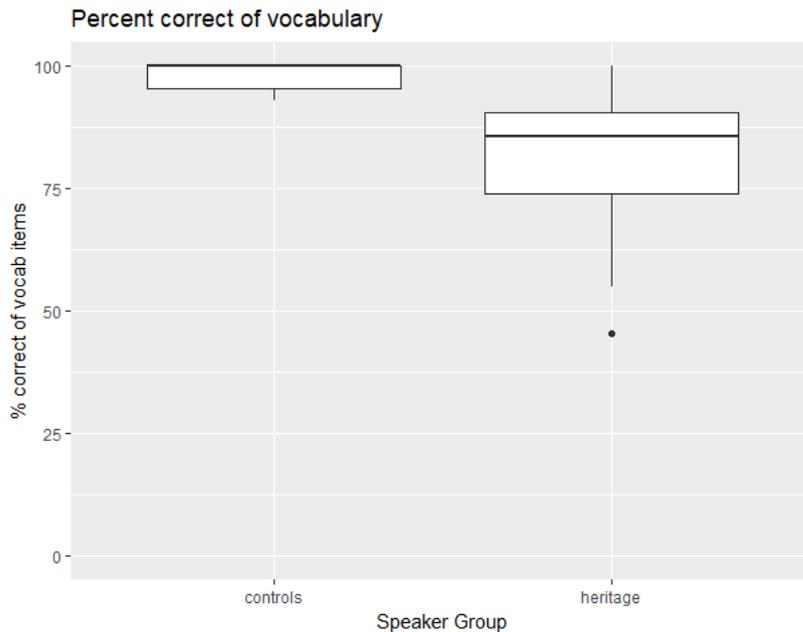


Figure 3.7: The percent of vocabulary items that participants identified using an appropriate label and article.

vided an incorrect noun. There is a question of the sensitivity of the task: it is possible that speakers may have guessed the gender of the word correctly, rather than knowing it. This is more a concern for the masculine nouns. Given that heritage speakers of Spanish (as well as children) tend to over-generalize the masculine, it is possible that participants may have used the masculine determiner and canonical ending *-o* on a noun as a guess, and their response would therefore not be an accurate reflection of their knowledge of the gender of the word. I cannot rule out the possibility that this is true, but given the frequency of the chosen lexical items in adult and child corpora of Spanish (Appendix A.1), we can have some confidence that the heritage participants were familiar with most of the lexical items, rather than guessing.

Polinsky (1997, 2006) and Godson (2003) further suggest that the time it takes to complete a lexical identification task can also serve as a proxy measure for syntactic proficiency. Therefore, the time it took participants to complete the lexical production task was recorded, and the results –

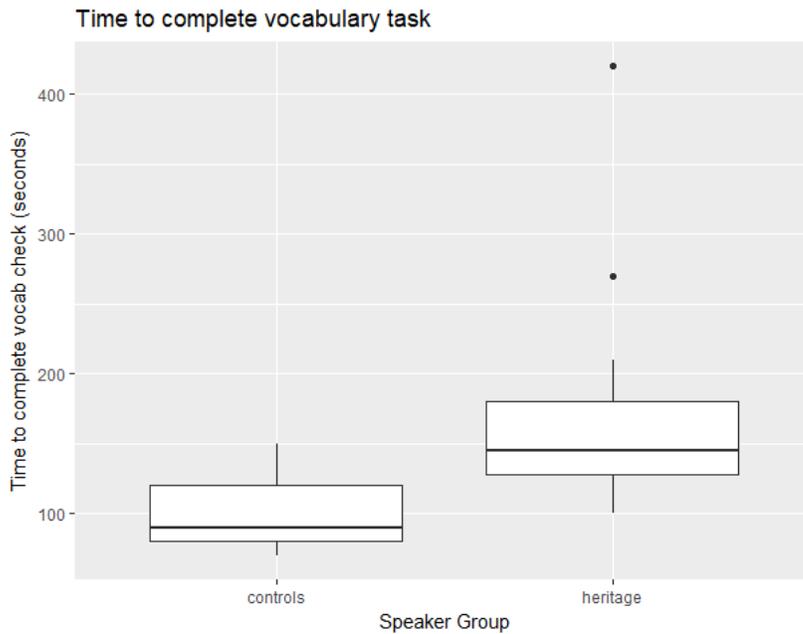


Figure 3.8: The time (in seconds) that it took participants to complete the lexical production task.

again split by group – are reported in Figure 3.8.

It should be noted that in this study this is a less reliable measure than the percentage of images correctly named. This is because participants varied in how much time they spent on a particular item that they could not recall a name for before giving up and moving to the next one. While some heritage speakers chose to think for quite a while, others would quite quickly skip forward to the next image if they didn't know it. Nevertheless, we do see a clear overall slow-down in the heritage participants, as compared to the controls, as we would expect. The heritage group on average took 164 seconds (standard deviation 72 seconds) to complete the task, while the control group on average took 102 seconds (standard deviation 27 seconds).

In addition to the proxy proficiency measures from the oral naming task, the LEAP-Q provides a set of self-reported proficiency measures. Figure 3.9 provides histograms that compare how heritage speakers and control speakers reported their proficiency in speaking Spanish, understanding spoken

Spanish, and reading Spanish. Participants rated their proficiency on a scale of 0 to 10, with 10 being fully proficient and 0 indicating no proficiency. While both groups' responses cluster on the upper end of the scale, the heritage speakers' responses are more distributed toward the middle of the scale in each of these measures. It should also be noted that heritage speakers are known to misrepresent their proficiency (Kang & Kim (2012) suggest their self-assessment often reflects the degree to which they identify with an ethnicity or culture rather than linguistic ability; see Chapter 6 of Montrul (2016), and Chapter 3 of Polinsky (2018) for further discussion), suggesting that there is a possibility that their self-reported scores on the LEAP-Q are inflated.

The LEAP-Q also asked participants to self-report how much they perceive an English accent in their own speech when speaking Spanish (on a scale of 0 to 10, with 0 indicating no accent and 10 indicating a pervasive accent). A follow-up question asked how often the participant is identified by others as a “non-native” speaker of Spanish (on a scale of 0 to 10, 0 indicating never, and 10 indicating always). These self-reported scores are also presented in Figure 3.9. Heritage speakers are clearly much more widely distributed on the scale. Among the controls, one participant responded 10 to both of these questions, although the way the survey was designed makes it quite plausible that the participant was inattentive to the labeling on the scales and thought 10 to be analogous to the 10-end of the scale in the proficiency questions (see Appendix B for reference).

As mentioned, heritage speaker self-reports are known to be not be fully reliable (see Chapter 6 of Montrul (2016), and Chapter 3 of Polinsky (2018) for discussion). Therefore, in the analysis in Section 3.7, the results of the oral production task will be used as the sole proxy measures for proficiency. In order to underscore the validity of using these proxy measures of proficiency, we can consider how they correlate with the self-reported proficiencies from the LEAP-Q. Figure 3.10 shows the relationship between performance on the lexical production task – in terms of number of vocabulary items (out of 42) that were correctly labeled by the speaker – and the self-reported measures. For the heritage speakers, there is a significant positive correlation between vocabulary proficiency

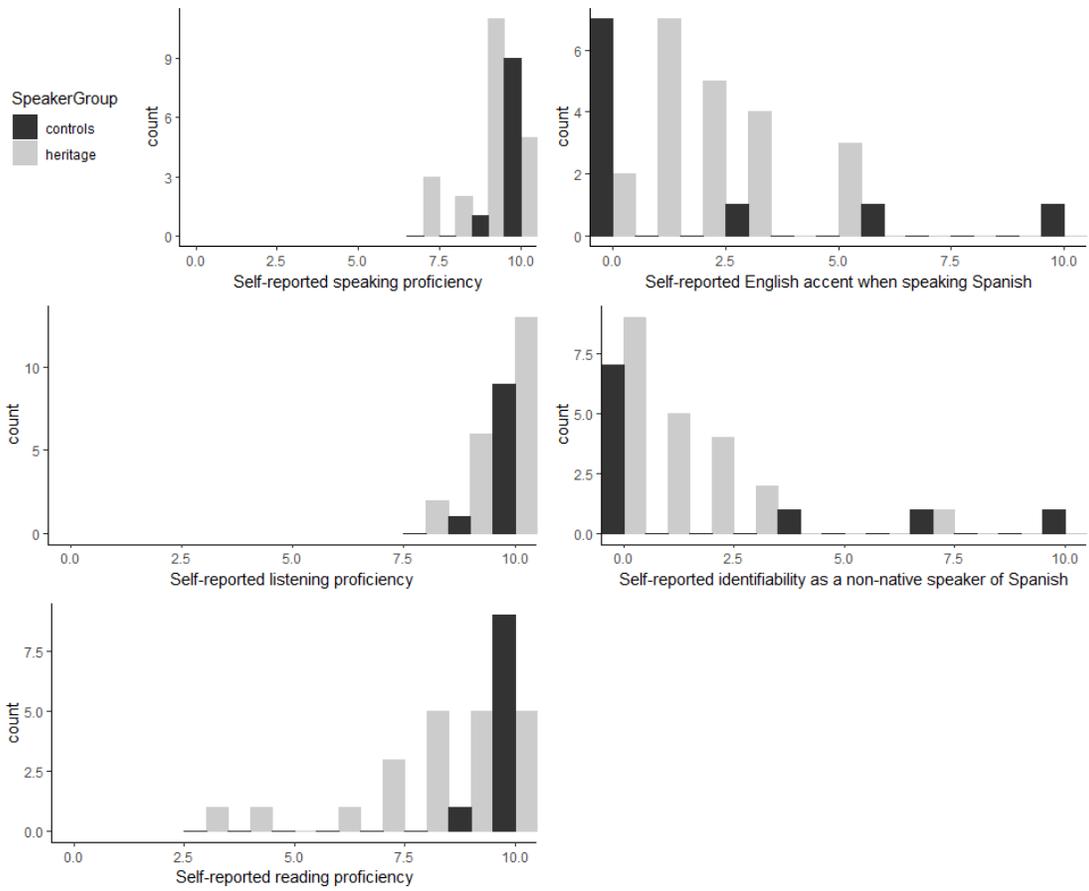


Figure 3.9: Responses of Spanish-speaking participants to the self-reported proficiency measure questions in the LEAP-Qs. All questions rated on a scale of 0 to 10.

and self-reported speaking proficiency ($R=0.65$, $p < 0.01$, Spearman method), self-reported listening proficiency ($R=0.43$, $p=0.05$, Spearman method), and self-reported reading proficiency ($R=0.56$, $p < 0.01$, Spearman method). The correlation is positive, because these variables were all coded such that the higher end of the scale should indicate higher proficiency. For the other two self-reported measures (self-perceived accent, and frequency of being identified as a “non-native” speaker of Spanish by others), the upper end of the scale corresponds to lower proficiency, i.e. having a strong accent and frequently being identified as a “non-native” speaker, and so we find negative correlations with the vocabulary task. The correlation with self-perceived English accent is significant ($R=-0.48$, $p < 0.05$, Spearman method), while the correlation with being identified as a “non-native” speaker is not ($R=-0.14$, $p=0.56$, Spearman method).

Figure 3.11 plots the relationship between speed on the lexical production task and the self-reported measures. Here the scatterplots suggest a negative relationship with self-reported proficiency in speaking, listening, and reading, and positive relationships with self-perceived English accent and being identified as a “non-native” speaker of Spanish, but none of these effects are significant.

What is evident from the range of self-reported proficiency measures and the lexical proficiency task is that although heritage speakers acquire Spanish as children, their language acquisition is non-convergent. That is, like L2 speakers, heritage speakers as a group show a range of proficiency and their language differs from that of monolingual adult speakers of the language. This is in-line with the discussion of the similarities and differences between L1, heritage, and L2 language acquisition in Chapter 2.

Given the possible unreliability of self-reported proficiency scores and the fact that lexical proficiency has previously been found to be a reliable proxy measure for syntactic proficiency, the correlations displayed in Figures 3.10 and 3.11 lend further credibility to using the scores from the oral picture-naming task as the primary measure of proficiency. This is the measure of proficiency that will be implemented in testing whether speaker proficiency plays a role in their reaction times in the

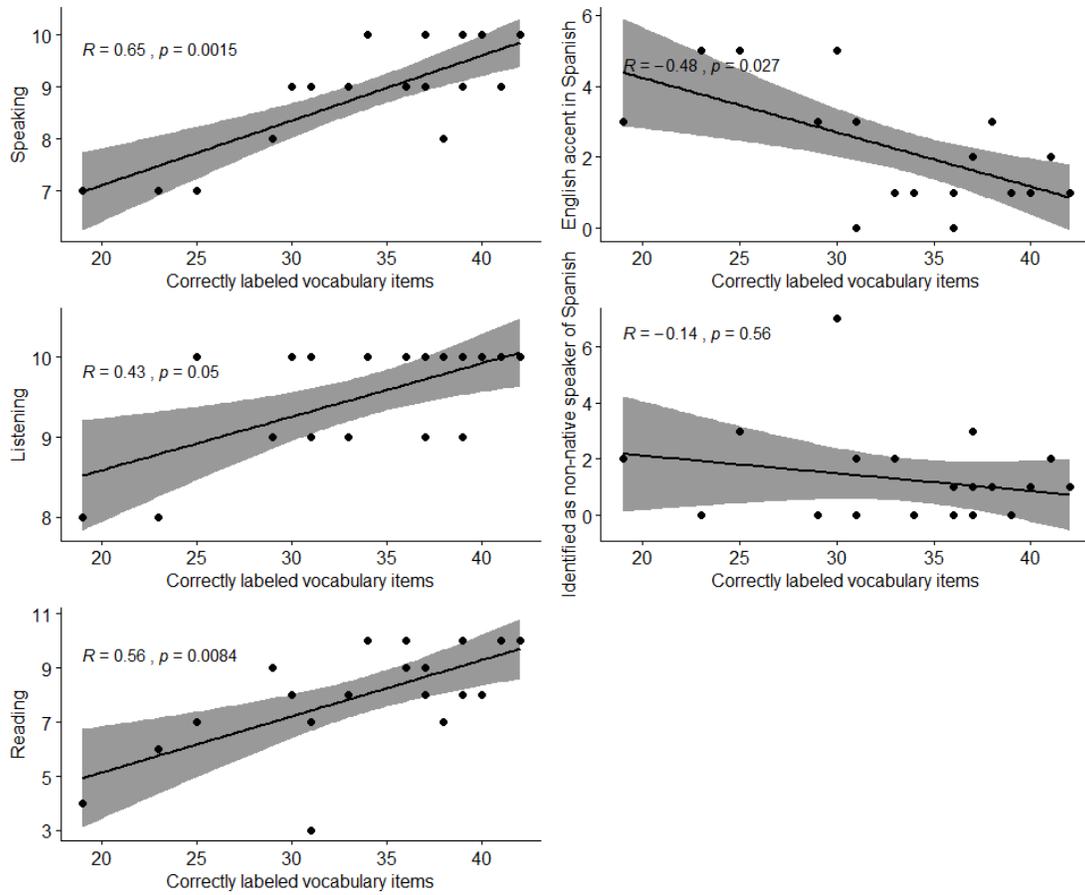


Figure 3.10: Relationship between performance on the lexical production task – in terms of number of vocabulary items (out of 42) that were correctly labeled by the speaker – and the self-reported measures.

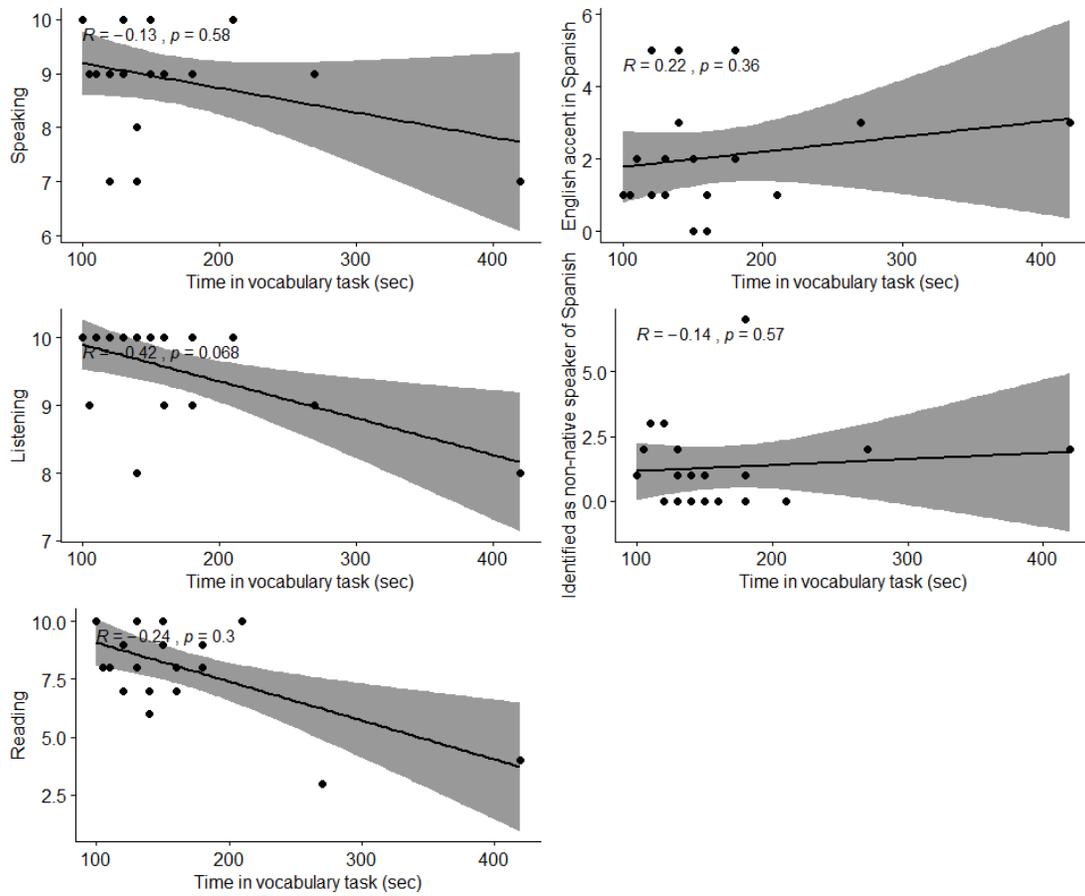


Figure 3.11: Relationship between speed on the lexical production task (in seconds) and the self-reported measures.

study in the next section.

3.7 ANALYSIS & RESULTS

The analysis began by removing those trials for which participants may not have had enough adequate information or knowledge to use gender information to facilitate processing of the target item. To do this, the lexical production responses were coded 0/1 in a slightly different way than when measuring proficiency. In this instance, a “correct” response was one that not only was an appropriate Spanish label for the item (with the correct corresponding gender), but it also matched the one anticipated by the experimental design. So if the participant provided a dialectal variant, this was considered incorrect for the purposes of identifying usable trials. Similarly, if a participant misidentified the referent of the image (ex. ‘eggs’ instead of the intended ‘nest’), this was also considered incorrect. With this coding, for every participant, I removed those trials for which the given participant had incorrectly labeled one or both of the images in the trial. For the control participants, this eliminated 134 data points, and for the heritage speakers this eliminated 898 data points. This is detailed in Table 3.3. The larger percentage of the data eliminated for the heritage speakers makes sense, given that they knew far fewer of the images than the control group did.

For the remaining trials, data was gathered from the Eyelink software and subsequently cleaned and analyzed in R. All data points were trimmed outside of two standard deviations of the mean based on all looking times. This left 824 data points for the controls (having eliminated 62) and 1192 data points for the heritage group (having eliminated 52). The SR Research Eyelink software records the time of fixation for up to three first fixations on an interest area. 17 trials were removed because

	Before trimming	After lexical task	After trimming
Controls	1020	886 (-13%)	824 (-7%)
Heritage	2142	1244 (-42%)	1192 (-4%)

Table 3.3: The number of data points for both of the speaker groups in the Spanish experiment, along with what percent of the data was removed at each stage of data cleaning: removal of unknown or different image labels from the lexical task, and trimming of reaction times outside of two standard deviations of the mean.

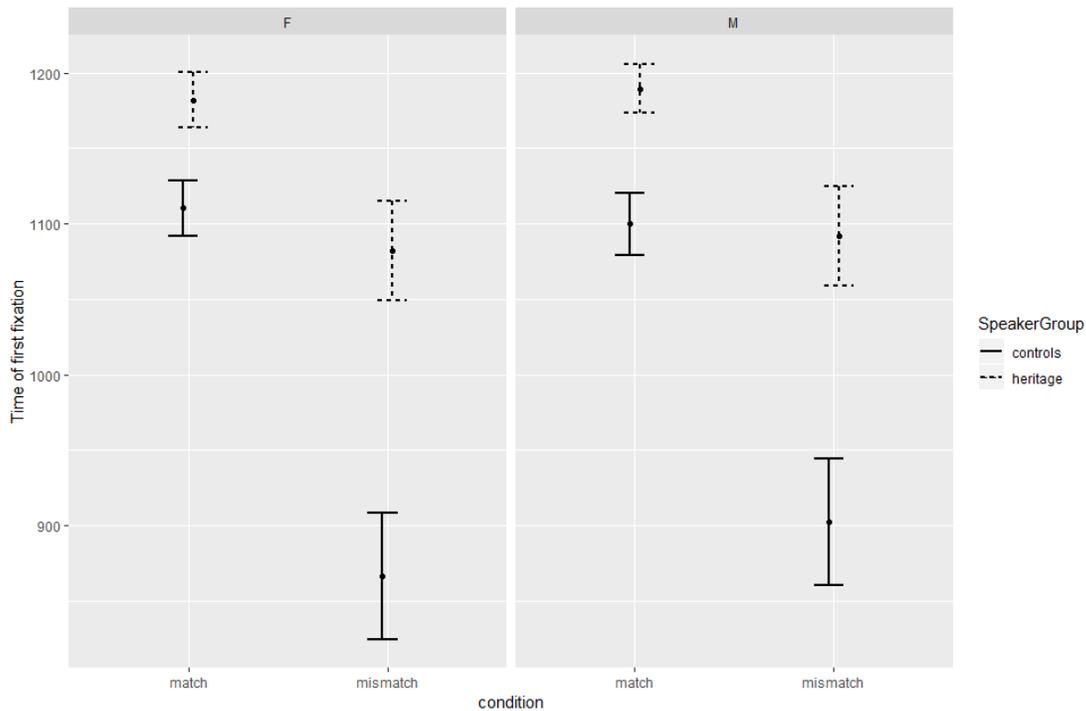


Figure 3.12: Mean time of first fixation on the target item (with 95% confidence intervals) across match and mismatch conditions for the different head-noun conditions, compared across the control and heritage groups.

the participants never fixated on the target items. For every remaining trial, the earliest fixation in the interest area later than 3250ms after the start of the trial (i.e. the onset of the article) was determined to be the time of first fixation on the target item. The time of first fixation past the critical period of the onset of the definite article was compared across gender match/mismatch conditions and across speaker groups. The data are plotted in Figure 3.12. Note that it takes 200 ms to launch a saccade, meaning that the 200 ms mark on the plot would be the earliest point at which a participant might fixate on an item.

I fit a linear mixed-effects model to the data, predicting time of first fixation by GROUP, CONDITION, and TRIAL, as well as their INTERACTIONS (pairwise, as well as the interaction of all three variables), with random intercept and slope for CONDITION conditioned on PARTIC-

	<i>Dependent variable:</i>		
	<i>Time of first fixation</i>		
group=heritage	82.453	(22.091)	***
condition=mismatch	-223.140	(30.834)	***
trial number	16.168	(8.314)	*
group : condition	127.493	(38.110)	***
group : trial number	-10.084	(10.973)	
condition : trial number	-53.089	(13.696)	***
group : condition : trial number	11.054	(17.845)	
constant	1,107.369	(17.768)	***
Observations	2,016		
Log Likelihood	-13,454.150		
Akaike Inf. Crit.	26,932.310		
Bayesian Inf. Crit.	26,999.610		
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01		

Table 3.4: Output of the mixed effects model fitted to the Spanish data for all conditions.

IPANT (Table 3.4). The model found a significant main effect of GROUP ($p < 0.01$) and CONDITION ($p < 0.01$), as well as a significant interaction between GROUP and CONDITION ($p < 0.01$) and a significant interaction between CONDITION and TRIAL ($p < 0.01$). Due to the presence of interaction effects we cannot interpret the main effects, so we will consider each of the interactions. In turn, a similar model with an added GENDER predictor found no significant effect of the gender of the target item, so data are not presented split according to target gender. For completeness, separate analyses for feminine target conditions and for masculine target conditions are provided in Appendix C.1 and Appendix C.2, respectively.

3.7.1 HERITAGE VS CONTROL GROUP RESULTS

First we interpret the interaction between GROUP and CONDITION. The control participants (mean = 1021 ms, sd = 233 ms) were overall faster than the heritage participants (mean = 1145 ms, sd = 215 ms), and both groups were faster on the mismatch trials (control mean = 1105 ms, sd = 267 ms; heritage mean = 1186 ms, sd = 164 ms) than on the match trials (control mean = 884 ms, sd = 157 ms; heritage mean = 1087 ms, sd = 261 ms). However, the difference between first fixation on match versus mismatch conditions differed between groups: There was a larger boost in speed on the mismatch conditions for the control participants than for the heritage participants.

We can probe this effect further by considering the by-participant results in Figure 3.13. The slopes for the control participants are overall much steeper than those of the heritage participants, who show more variation. While some of the heritage participants appear to perform in a control-like manner, most of them have decidedly flatter slopes, indicating a smaller difference between time of first fixation on match versus mismatch conditions. Notably, however, the heritage speakers (with one exception) do consistently perform faster on the mismatch trials than on the match trials.

To confirm that the effect of condition is non-zero for the heritage population, I ran a separate linear mixed effects model restricted to the heritage speaker data, predicting time of first fixation by CONDITION, and TRIAL, as well as their INTERACTION, with random intercept and slope for CONDITION conditioned on PARTICIPANT. The model found a significant main effect of CONDITION as well as the interaction of CONDITION and TRIAL NUMBER. The effect of condition shows that heritage speakers overall performed faster on mismatch trials than on match trials. This interaction is illustrated in Figure 3.15 and discussed in Section 3.7.4.

Recall the model interpretations from above:

- (16) 1: If heritage speakers can use gender information to facilitate lexical retrieval, then the time of first fixation on target items will be faster in mismatch conditions than in match condi-

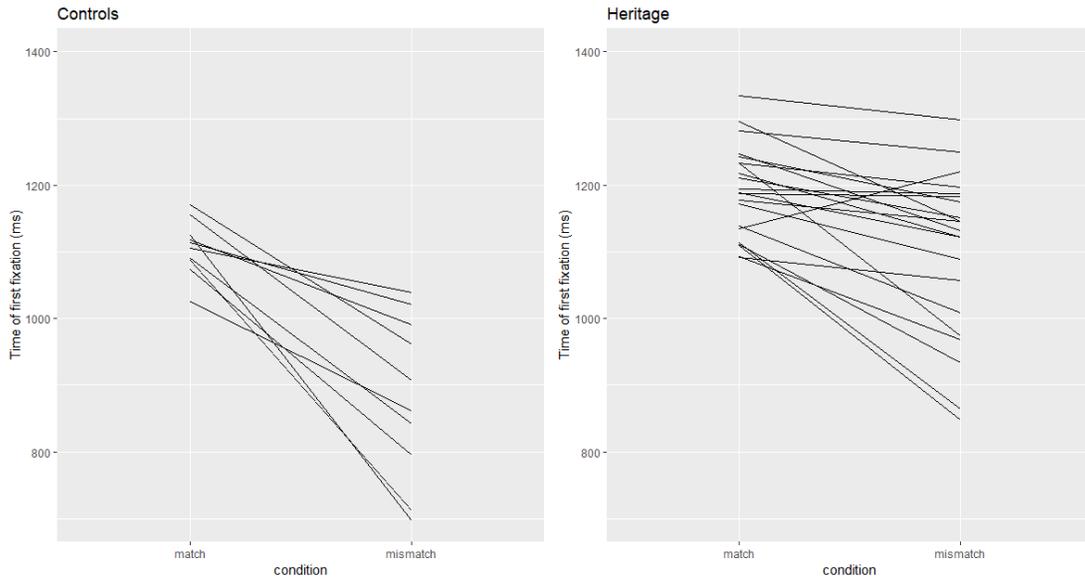


Figure 3.13: By-participant average time of first fixation on match versus mismatch conditions, split across participant groups.

<i>Dependent variable:</i>			
<i>Time of first fixation</i>			
condition=mismatch	-97.624	(19.200)	***
trial number	6.376	(7.089)	
condition : trial number	-42.193	(11.326)	***
constant	1,190.942	(14.265)	***
Observations	1,192		
Log Likelihood	-7,953.762		
Akaike Inf. Crit.	15,923.520		
Bayesian Inf. Crit.	15,964.190		

Note: *p<0.1; **p<0.05; ***p<0.01

Table 3.5: Output of the mixed effects model fitted to the data for the heritage group for all conditions.

tions.

2: If heritage speakers cannot use gender information to facilitate lexical retrieval, then the time of first fixation on target items will be about equal in match and mismatch conditions.

The results here are in line with (1), suggesting heritage speakers can use gender to facilitate spoken word recognition.

3.7.2 PROFICIENCY VARIABLES

In order to determine whether proficiency of the heritage speakers predicts their ability to use gender to anticipate the target item, I considered only the subset of the data that came from heritage speakers and fit mixed-effects linear models that included these proficiency measures.

The first model predicted FIXATION by CONDITION, TRIAL, and VOCABSUM, as well as their interaction. The model included a random intercept and slope for CONDITION, conditioned on PARTICIPANT. The VOCABSUM variable indicated the number of vocabulary items that a participant was able to give an appropriate label for in Spanish. The model did not find a significant effect of VOCABSUM.

The second model predicted FIXATION by CONDITION, TRIAL, and VOCABTIME, as well as their interaction, and included random intercept and slope for CONDITION, conditioned on PARTICIPANT. Here, VOCABTIME indicates the amount of time participants took to complete the vocab task, in seconds. The model did not find a significant effect of VOCABTIME.

3.7.3 DEMOGRAPHIC VARIABLES

In order to determine whether any of the demographic variables influence whether heritage speakers are able to use gender predictively and the speed at which they do so, I fit linear mixed-effect models like the ones above, but adding each of the demographic variables and their interaction with CON-

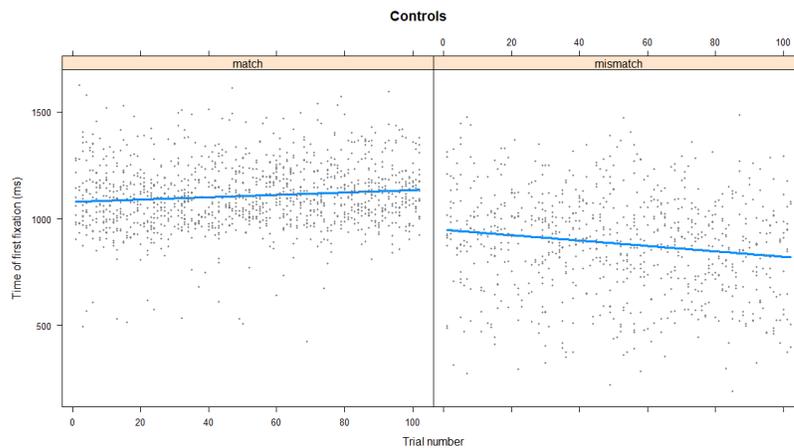


Figure 3.14: The average time of first fixation for control participants over the duration of the task, indicated by trial number (1-102).

DITION and TRIAL as fixed effects. None yielded significant effects. The variables tested were level of education, number of years spent living in a Spanish-speaking country, number of years spent living in a Spanish-speaking family, number of years in a Spanish-speaking work or school environment, and age of the heritage speaker at the time of participation.

3.7.4 EFFECT OF TRIAL NUMBER

We can now turn to consider the interaction of CONDITION and TRIAL. Overall, participants were faster on mismatch conditions than on match conditions, but this effect changed over the duration of the task. The effect is represented visually for control participants in Figure 3.14 and for heritage participants in Figure 3.15.

Figures 3.14 and 3.15 show a clear learning effect in the mismatch trials, as participants increase in speed over the course of the task. For the match condition, we do not have this effect. In fact, there appears to be a trend for slower times of first fixation in later trials (visually more pronounced for control participants, although this is not significant), possibly indicative of a tiredness effect.

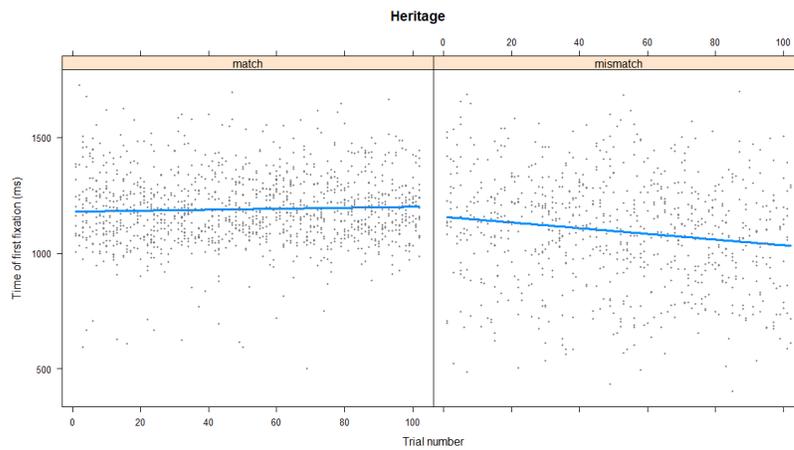


Figure 3.15: The average time of first fixation for heritage participants over the duration of the task, indicated by trial number (1-102).

3.8 DISCUSSION

The data in Section 3.7 suggest several key results. First, control speakers were significantly faster on mismatch trials than on match trials, replicating the (very robust) previous findings on the ability of control participants to use gender to facilitate lexical retrieval in online processing. Second, the heritage speakers were also consistently faster to fixate on the target item in mismatch condition trials than in match condition trials. This indicates that heritage speakers are able to use gender information on articles to facilitate lexical retrieval in online processing in a control-like manner. However, we did find that heritage speakers were overall slower in the task than the control group (similar to the children in [Lew-Williams & Fernald \(2007\)](#)), and that the increase of speed on mismatch conditions was less pronounced for heritage speakers than for the control group.

3.8.1 PREVIOUS PREDICTIONS

Recall the two competing hypotheses regarding the underlying cause for differences between L1 adults and children versus L2 adults:

(17) Hypothesis 1: If the ability to use gender to facilitate word recognition is associated with the nature of the learning experience – the so-called “tight link” – then heritage speakers should pattern with L1 adults and children in being able to use gender information on determiners to anticipate the subsequent nouns.

Hypothesis 2: If this ability is instead tied to language proficiency, then heritage speakers should pattern with L2 learners who, in the [Lew-Williams & Fernald \(2010\)](#) and [Grüter et al. \(2012\)](#) experimental setup, were unable to use gender on determiners to anticipate subsequent nouns.

The heritage speakers in this study were able to use gender information on determiners to facilitate spoken word recognition in a monolingual-like manner, offering support for the “tight link” hypothesis (Hypothesis 1). Heritage speakers are like L1 adults and children in that the input to their grammar is early and naturalistic. They learn Spanish by parsing the speech stream, and their early grammar likely goes through the same stages of parsing article + noun units as chunks, only subsequently recognizing independent morphemes. Thus an important possible implication arises: although heritage speakers of Spanish are overall slower in processing Spanish, they still have an advantage over L2s in the ways in which they can use language information in online tasks, as previously suggested by [Montrul et al. \(2014\)](#) but now with supporting evidence from a more sensitive measure.

It should be noted that the ability of L2 speakers to use gender in online tasks in their L2 has been intensely studied, and recent findings suggest it may be a “graded phenomenon” ([Curcic et al. 2018:49](#)). Factors that have been investigated in relation to this ability include overlap between the L1 and the L2 ([Dussias et al. \(2013\)](#), to be discussed in detail below), as well as proficiency, which was previously discussed ([Hopp \(2013, 2016\)](#)). Recently, the question of awareness has also been considered. [Curcic et al. \(2018\)](#) conducted a study in which L1 speakers of Dutch learned a fake language

that had a two-gender system with a clear morphophonological cue to gender on noun endings, and with pronominal determiners that varied with gender. After a training period, they conducted an eye-tracking study similar in experimental design as the studies in [Lew-Williams & Fernald \(2007, 2010\)](#), [Grüter et al. \(2012\)](#). They found that participant ability to use gender in the fake language was conditioned on whether participants were overtly aware of the pattern of agreement between determiners and nouns, and whether they were aware that determiners were useful in the eye-tracking portion of the study. Awareness was determined through a post-experimental debriefing.

The question of proficiency for heritage speakers was addressed in the analysis, with the results reported in Section 3.7.2, to be discussed below. The results of this study cannot bear on whether gender in the dominant language impacts heritage speaker performance on the task, but this will also be discussed further below. Left outstanding is the issue of awareness. This was not data that was gathered in the course of this experiment, so I cannot test to what extent this played a role in the study. However, it is generally understood that heritage speakers have low metalinguistic awareness of their language due to a lack of formal schooling in the heritage language, per the discussion in Chapter 2. Therefore, had this study included an assessment of awareness, I imagine there would have been a mix of responses. Future iterations of this study may fold this into the experimental procedure.

3.8.2 PROFICIENCY

It is also notable that the study did not reveal any correlation between demographic/proficiency variables and the heritage speakers' speed in fixating on the target item in the task. As will be discussed, null results are difficult to interpret in a frequentist framework, particularly with small sample sizes (Section 5.1). There are many more fine-grained variables that could be extracted from the proficiency scores that might prove to be correlated with the results after all. One might, for instance, determine for each lexical item how long it took a participant to access the lexicon and

name the item, and test this for correlation with the time to fixation on trials with the given target item. However, should the lack of correlation hold up to further statistical scrutiny, this would be evidence that more fine-grained methodologies such as eye-tracking allow us to identify a type of inter-heritage-speaker variation that is not attributable to external factors but rather is an inherent property of the heritage speaker population.

3.8.3 LACK OF GENDER ASYMMETRY

As noted in Section 3.7 and as detailed in the models and plots provided in Appendix C, this study found no asymmetries between heritage or monolingual ability to use masculine or feminine gender to facilitate lexical retrieval. This is in-line with the previous findings for monolingual Spanish-speaking adults and children (Lew-Williams & Fernald (2007, 2010), Grüter et al. (2012), Dussias et al. (2013)). It is also in-line with the findings on L2 speakers of Spanish whose L1 is English (Grüter et al. (2012), Dussias et al. (2013)). However, it contrasts with the findings in Dussias et al. (2013) for L2 speakers of Spanish whose L1 is Italian. The results of their study suggest that L1-Italian, L2-Spanish participants are only able to use feminine gender in Spanish to facilitate lexical retrieval.

Dussias et al. (2013) argue that the asymmetrical finding is due to the nature of the definite articles in Italian, the L1 of the L2 speakers participating in their study. Given that L2 language acquisition is different by some metrics from heritage language acquisition, it is not surprising that the heritage speakers differ from L2 speakers in this respect. However, these findings do raise the question of how Spanish heritage speakers' ability to use gender in an online task would be impacted by a dominant language with a grammatical gender system. The heritage speakers studied here have English as their dominant language, and English for the most part does not have gender. In a follow-up study conducted in heritage speakers living in Germany, I compare the present findings to how Spanish heritage speakers with German as the dominant language use grammatical gender in online tasks.

German has three genders – masculine, feminine, and neuter – and given that it belongs to the Germanic language family, its gender system shows less overlap with Spanish than the Italian system does.

While the asymmetrical gender findings in [Dussias et al. \(2013\)](#) pertain only to L2 speakers, asymmetries between genders have been noted for monolingual speakers. [Loerts \(2012\)](#), [Loerts et al. \(2013\)](#), and [Brouwer et al. \(2017\)](#) report asymmetrical findings for Dutch, which has two genders – common and neuter. Gender in Dutch is encoded in the definite article *de* for common gender and *het* for neuter gender, with few morphophonological cues to gender on nouns. The genders are unequally distributed in the lexicon, with the common gender accounting for 75% of the lexicon. The system is also notable for the fact that Dutch children do not appear to fully acquire the system until around age 6 (see discussion in [Brouwer et al. \(2017\)](#)).

The difficulty in interpreting the asymmetrical findings for Dutch gender in eye-tracking is that the results are conflicting. Whereas [Loerts \(2012\)](#) and [Loerts et al. \(2013\)](#) find that Dutch speakers can only use common gender to facilitate lexical retrieval, whereas [Brouwer et al. \(2017\)](#) find the opposite, that Dutch adults and children can only use the neuter gender. Various arguments are put forth in support of one or the other argument, including a discussion of where *de* and *het* appear elsewhere in the language, making them less useful in prediction: *het* is also used for diminutives as well as personal or indefinite pronouns, whereas *de* (common singular) is also used for neuter plural nouns.

Nevertheless, the results in this study do not find a gender asymmetry either for the monolingual adults or the heritage speakers. This is expected, given the lack of gender effects for monolingual adults and children, as well as L1-English L2-Spanish speakers, reported elsewhere in the literature. In the next chapter, the analysis of Polish gender will reveal gender asymmetries in how the presence of a particular gender as a distractor impacts participants' ability to use a given target gender.

3.9 GOING FURTHER

The results presented in this study indicate that heritage speakers are monolingual-like in their ability to use grammatical gender information to facilitate spoken word recognition, with the caveat that they are slower than monolinguals in doing this. To gain a better understanding of why this is, we should consider in more detail what linguistic knowledge or ability this linguistic processing really reflects. As [Lew-Williams & Fernald \(2007\)](#) discuss, there are a few possibilities. First, this could be a surface phenomenon, tapping into speakers' attention to probabilistic properties of their language. The idea here is that transition probabilities between an article of a particular gender and candidate nouns guide the participant to orient to the target item faster in mismatch conditions ([Mintz \(2003\)](#), [Pine & Lieven \(1997\)](#)). Alternatively, this processing phenomenon may actually reflect speakers' abilities to access gender features in the grammar. [Lew-Williams & Fernald \(2007\)](#) explain: "A grammatical account posits that rule-based syntactic knowledge accounts for faster processing of nouns preceded by gender-marked articles ([Friederici & Jacobsen \(1999\)](#), [Lev-elt et al. \(1999\)](#)). Under this account, the article *la* primes all feminine nouns, leading to more rapid recognition of *pelota*."

We can begin to probe the nature of this process further by considering what factors affect heritage speakers' ability to use gender information to facilitate lexical retrieval. In other words, what variables might make heritage speakers perform more or less like monolinguals? Answering this question may guide us to identifying the process or knowledge that speakers are using in performing these eye-tracking tasks, while simultaneously considering what factors play a role in shaping heritage speakers' linguistic ability.

A few intuitive candidates for this investigation present themselves. First, the complexity of the gender feature itself may play a role. Spanish has two grammatical genders, but what if the language has three genders or more? Does a more complex gender feature put too much stress on the her-

itage speaker's processing abilities? These questions will be discussed further in the next chapter, in which the present study is replicated on monolingual and heritage speakers of Polish. The results in Chapter 4 will also motivate asking whether it is the nature of the article (versus other elements in the nominal domain) – and how gender information is accessed and stored on the article – that impacts speakers' ability to use the information in online processing. In Chapter 5, comparing the results from heritage speakers of Spanish and Polish, I will suggest directions for further study to investigate these questions. Transfer from the dominant language is another intuitive candidate for further investigation, and a follow-up study conducted in Germany in February 2019 will address this question.

On a final note, a possible extension of this study might bear on the question of the relationship between number and gender in the heritage grammar. Recall that [Fuchs et al. \(2015\)](#) analyze number and gender as represented independently in the monolingual grammar, but in a follow-up study determine that number and gender are bundled in the heritage grammar ([Scontras et al. \(2018\)](#)). They predict that therefore number and gender information should be more opaque for heritage speakers, with the effect of being more difficult to access in online processing. The study presented here used only singular articles; given the assumption that the singular number value in Spanish is both morphologically and representationally underspecified, this implies that only the gender feature was located on the articles in the auditory stimuli. In an extension, we might incorporate plural articles *los* and *las* to determine how the presense of both number and gender features affects participants' ability to use either feature in an online task. The predictions made by [Scontras et al. \(2018\)](#) set an expectation that heritage speaker performance on the task should be impacted, whereas monolingual speaker performance need not be (or both may be impacted, with a greater effect for the heritage group than the control group). The results of this extension would only be interpretable as bearing on the relationship between syntactic features if it can be convincingly argued that participants access the syntax and its features in this task, a question that will be addressed

further in Chapter 5.

4

Eye-tracking study: Heritage Polish

WE TURN TO POLISH, to test whether a more complex gender feature impacts heritage speaker ability to use grammatical gender to facilitate lexical retrieval. In Chapter 3 we saw that heritage speakers of Spanish are able to use grammatical gender in a monolingual-like manner, matching both adults and children in their ability to do this. In order to probe further the nature of the mechanism that allows heritage speakers to do this, I aim to test what properties or conditions impact heritage

speaker performance on the task. The first candidate property is the complexity of the gender feature itself: Does having a three-gender system prove more or less difficult for heritage speakers to maintain and to use in online tasks? Are there any asymmetries in their abilities to use some subset of gender values but not the others in an anticipatory manner?

Polish has three genders and a readily accessible community of heritage speakers that provides a testing ground for this question. Additionally, since the methodology and research questions have not been conducted on monolingual speakers of Polish either, this chapter contributes novel data to the broader subfield of anticipatory use of gender in eye-tracking studies. In fact, to my knowledge, this is the first psycholinguistic study of grammatical gender in Polish more broadly. Comparing the results of the monolingual speakers in this study with those of the heritage speakers, I will ask (a) whether heritage speakers of Polish are able to use gender to facilitate lexical retrieval, and (b) if there are any asymmetries in the results that shed light on the representation of gender in the heritage and/or monolingual grammar.

There is reason to think the impact of a more complex gender feature could go either way. The notion that a three-gender system may be more difficult for heritage speakers to use to facilitate lexical retrieval or may be subject to asymmetries stems from work on gender in Heritage Russian. Baseline Russian has three genders – masculine, feminine, and neuter, but Polinsky (2008c) demonstrated that some heritage speakers reanalyze the gender system. In the reanalyzed gender of this speaker group, the neuter is subsumed under the feminine, leaving only a masculine/feminine distinction. Polinsky motivates the loss of the neuter in the following way: "The overall reanalysis of the baseline noun categorization system by heritage speakers targets those classes that pose a challenge to L1 learners and/or are numerically smaller [...] and assimilates nouns of those classes to stronger 'attractors'" (Polinsky 2008c:63). While a similar study has not been conducted on heritage speakers of Polish, this suggests that the three-gender system in Polish might also be susceptible to such reanalysis. Certainly Janssen (2016) shows that neuter is more difficult in the acquisition of Pol-

ish gender, and Brehmer & Rothweiler (2012) present data that bilingual children do not produce correct neuter agreement, instead overgeneralizing the masculine (discussed further in Section 4.1.2). Taken together, this points to the possibility that a three-gender system may be more difficult for heritage speakers to use in online tasks, and that asymmetries may be present where the neuter is involved.

This need not be the case, however. A contrasting line of thinking would suggest that gender in Polish may be easier for heritage speakers to use in online tasks, precisely because it is more complex. This notion underlies some discussions of the acquisition of grammatical gender. For instance, Unsworth et al. (2014) conducted an elicitation study on bilingual children with Greek as the heritage language and English as the dominant language, as well as children with Dutch as the heritage language and English as the dominant language. Dutch has two genders (common and neuter), whereas Greek has three (masculine, feminine, and neuter). Despite the relatively higher complexity of the Greek gender system, Unsworth et al. (2014) found that the Greek-English bilingual children were more target-like in their production of gender assignment and agreement than the Dutch-English bilingual children, who only had to learn two genders. The authors argue that a gender system with more features is more salient, and that in Greek the gender system is more transparently reflected in the morphophonology of nouns and determiners. Thus Greek bilingual children are quicker to learn the gender system in their heritage language than Dutch bilingual children, whose gender system is less salient or transparent. Additionally, Unsworth et al. (2014), as well as Unsworth et al. (2011) based on a similar comparison of bilingual acquisition of Greek voice and Dutch gender, argue that the age of onset of bilingualism is not a determining factor in bilingual acquisition of these gender systems, but that the amount of input may play a role. The argumentation would proceed similarly for adult heritage speakers. If they are more likely to learn a more complex gender system earlier and more completely due to its salience and transparency, then this may translate to their ability to maintain this gender system over time and be able to use it in online

processing.

This chapter is structured as follows. Section 4.1 provides the necessary background on the formal and acquisitional properties of Gender in Polish. Section 4.2 discusses the experimental design and how it differs from that of the Spanish study. Section 4.3 discusses the model interpretations relevant here, and in Section 4.4, I discuss the participants and some demographic factors. Section 4.5 walks through the procedure implemented before Section 4.6 gives some measures of proficiency of the participants. The results are presented in Section 4.7 and the implications of these findings are discussed in Section 4.8.

4.1 BACKGROUND

4.1.1 GENDER IN POLISH

Polish is generally recognized as having three gender: masculine, feminine, and neuter. As discussed previously, the most reliable indicators of gender are the forms of associated words, as illustrated for demonstratives and adjectives in (1-a) for masculine, (1-b) for feminine, and (1-c) for neuter.

- (1) a. Ten star-y wazon był w kuchni.
DEM.M.SG old-M.SG vase.M be.PST.3SG.M in kitchen
'That old vase was in the kitchen.
- b. Ta star-a książka była w kuchni.
DEM.F.SG old-F.SG book.F be.PST.3SG.F in kitchen
'That old book was in the kitchen.
- c. To star-e wiadro było w kuchni.
DEM.N.SG old-N.SG bucket.N be.PST.3SG.N in kitchen
'That old bucket was in the kitchen.

As in Spanish, there are morphological correlates of each of the genders. The feminine gender is typically associated with the ending *-a* (2-a), with many exceptions (2-b). Likewise, the neuter is

typically associated with the endings *-o*, *-e*, or *ę* (3-a), again with some exceptions (3-b).

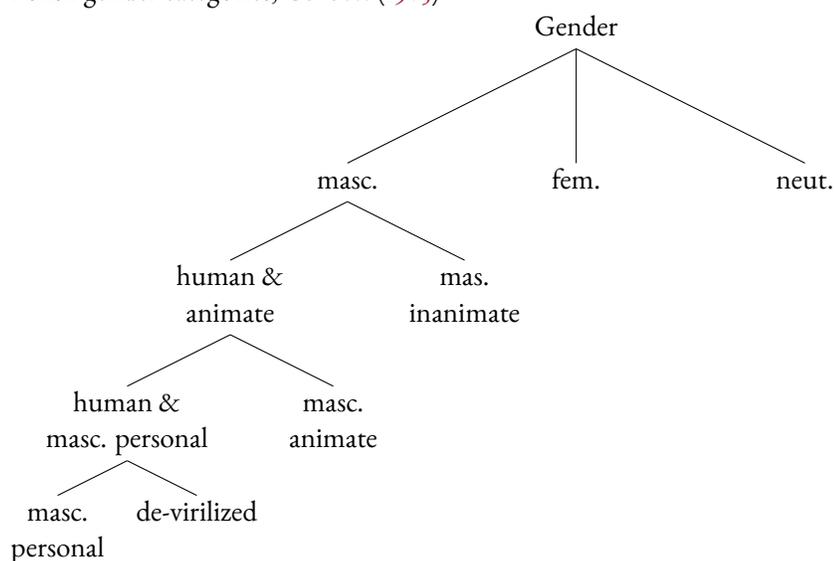
- (2) a. *ta dziewczyna* ‘this girl, fem.’
ta kobieta ‘this woman, fem.’
ta koszula ‘this shirt, fem.’
ta książka ‘this book, fem.’
- b. *ta sól* ‘this salt, fem.’
ta pięść ‘this fist, fem.’
ta rzecz ‘this thing, fem.’
- (3) a. *to jajko* ‘this egg, neut.’
to okno ‘this window, neut.’
to wiadro ‘this bucket, neut.’
to dziewczę ‘this maiden, neut.’
- b. *to coś* ‘this something, neut.’
to menu ‘this menu, neut.’

Most masculine words in Polish end in consonants (4-a), although some – typically relating to professions – can end in *-a* (4-b).

- (4) a. *ten chłopiec* ‘this boy, masc.’
ten pies ‘this dog, masc.’
ten stół ‘this table, masc.’
ten wazon ‘this vase, masc.’
- b. *ten mężczyzna* ‘this man, masc.’
ten artysta ‘this artist, masc.’
ten sędzia ‘this judge, masc.’

The actual number of gender values in Polish has been largely debated in the formal literature. The most broadly accepted analysis and the one I adopt here is Corbett (1983), who argues for three genders (masculine, feminine, and neuter), with several *subgenders* (Wertz (1977), Corbett (1983), Brown & Hippisley (2012), among others). Corbett (1983) states: “Subgenders are agreement classes which control minimally different sets of agreements [...] typically [...] not including the most basic form (usually the nominative singular).” He adds that a qualification for being considered a subgender is “consistent agreement patterns” (Corbett 1983:7). The gender categories proposed for Polish by Corbett (1983) are presented in (5). Notably, both Corbett and the competing accounts of Polish gender posit a flat structure between masculine, feminine, and neuter. This will be addressed in Section 4.8.2.*

(5) Polish gender categories, Corbett (1983)



*In alternative accounts to Polish gender that posit only three subgenders of masculine, split between masculine personal and masculine de-virilized that is omitted. De-virilized nouns are referentially masculine personal nouns with a pejorative connotation, such as *brudas* ‘dirty person’. Their declension patterns are largely those of the masculine personal, with a handful of exceptions (Zagorska-Brooks & Nalibow (1970), Brown & Hippisley (2012)).

	singular		plural			
	animate	inanimate	personal	de-virilized	animate	inanimate
nom.	duży	duży	duzi	duże	duże	duże
gen.	dużego	dużego	dużych	dużych	dużych	dużych
dat.	dużemu	dużemu	dużym	dużym	dużym	dużym
acc.	dużego	duży	dużych	dużych	duże	duże
inst.	dużym	dużym	dużymi	dużymi	dużymi	dużymi
loc.	dużym	dużym	dużych	dużych	dużych	dużych

Table 4.1: The declensional paradigm of the Polish adjective *duży* 'big' in the masculine subgenres. In the singular, the only distinction is between animates and inanimates. In the plural, there is evidence for four subgenres.

Evidence for the subgenres of masculine in Polish following [Corbett \(1983\)](#) comes from adjectival declensional paradigms beyond the nominative singular, as illustrated in Table 4.1. The split between the masculine animate and the masculine inanimate is evident in the accusative singular, where adjectives agreeing with masculine animate nouns take the suffix *-ego* whereas adjectives agreeing with masculine inanimate nouns take the suffix *-y*. Further subdivisions of the masculine animates can be seen in the plural morphology. Human animate nouns split from non-human animate masculine nouns in the accusative plural, while the division within the human animates into personal and de-virilized is evident in the nominative plural.

There are many competing, though less broadly accepted accounts. The number of genders in these analyses ranges from three ([Klemensiewicz 1960:51-52](#)) to nine ([Schenker \(1955\)](#)). Since gender is “reflected in the behavior of associated words” ([Hockett 1958:231](#)), these analyses largely focus on the inflectional patterns of nominal modifiers. The approaches vary in which class(es) of modifiers and which cases they target. Nominal modifiers that inflect to match the head noun in Polish include demonstrative pronouns, attributive and predicate adjectives, and relative pronouns, and verbs in the past tense and future imperfective also inflect for gender ([Swan \(2015\)](#)). Numerals are not classic agreement targets and are often treated separately from other Polish modifiers (see [Klock-](#)

mann (2012) for discussion). The nominal system has six nominal cases[†]: nominative, accusative, genitive, dative, instrumental, and locative. Case, number, and gender are fused in a single suffix (which can be phonologically null) both on nouns and on their modifiers.

The gender system recognized in the grammar of Polish *Gramatyka współczesnego języka polskiego* (Dukiewicz et al. (1984)) recognizes five Polish genders (masculine personal, masculine animate, masculine inanimate, feminine, neuter); this is based on accusative case markers on the adjective *dobry, dobra, dobre* ‘good’, first argued for by Mańczak (1956). In contrast, Szober (1963) considers the nominative case and identifies (the traditional) three Polish genders in the singular (masculine, feminine, and neuter) and two in the plural (masculine personal and other). This is the system typically taught in Polish schools Swan (2015). Swan (2015) identifies four Polish genders: masculine, feminine, neuter, and personal. Schenker (1955)’s system of nine genders is based on the inflection of the numeral *dwa* ‘two’; Saloni & Świdziński (2001) arrives at the same number of genders based on a broader set of numerals, but neither analysis has been received favorably by other researchers (see Swan (2015) for discussion).

As previously stated, I adopt here Corbett (1983) analysis, although following acquisition work by Krajewski (2005) and Brehmer & Rothweiler (2012), this chapter will focus solely on the broad distinction between masculine, feminine, and neuter. The various subdivisions of the masculine are based on animacy/virility and linked with different declension classes (i.e. different case paradigms), but in the nominative singular they collapse, and only the main three gender categories can be distinguished. As the discussion in Section 4.2 will reveal, the stimuli were designed in such a way as to only include inanimate stimuli, and the carrier phrase is designed so that the target lexical item occurs in the nominative case. This way, only the broad masculine/feminine/neuter distinction is relevant to the study.

[†]The vocative is not included here, as it does not form part of the syntax, and modern Polish speakers often use nominative forms in its place (cf. Luczynski (2002)).

DEFAULT GENDER IN POLISH

To determine what is the default gender agreement, we can consider (6), in which there is no overt controller. In the absence of the controller, the verb appears to be in the neuter. Similarly in (7), nouns with no referent or an unknown referent (the absence or underspecification of gender) also take neuter agreement.

- (6) Było zimno.
be.PST.3SG.N cold.N.SG
'(It) was cold (outside).'
- (7) a. To coś mnie denerwuje.
DEM.N.SG something ISG.DAT annoy.PRS.3SG.N
'This something is annoying me.'
- b. Nic się nie pokazało.
nothing REFL NEG appear.PST.3SG.N
'Nothing appeared.'

Using modifiers of coordinated NPs is not a reliable test for default gender agreement since Polish adopts the strategy of closest conjunct agreement ((8), adapted from [Willim \(2012\)](#)). Verbal agreement with a coordinated NP is uninformative because it is always plural, and plural morphology collapses the genders.

- (8) a. mały/*mała/*małe kot i mysz
small.M.SG/small.F.SG/small.PL cat.M and mouse.F
'small cat and mouse'
- b. mała/*mały/*małe mysz i kot
small.F.SG/small.M.SG/small.PL mouse.F and cat.M
'small mouse and cat'

The only distinction we see in verbal agreement in the plural is between the masculine personal (-i) and all others (-y). In coordinated NPs in which both NPs are personal but one is masculine and

the other feminine, the verb shows plural masculine personal agreement, as in (9-b), in contrast with (9-a). Thus only with animate personal nouns do we have enough information to posit a default for coordinated NPs: the masculine is the default for sex-differentiable nouns (in line with what [Kramer \(2015\)](#) suggests for Russian).

- (9) a. Basia i Zuzia poszły do sklepu.
 Basia.F and Zuzia.F go.PST.3PL to store
 ‘Basia and Zuzia went to the store.’
- b. Tomek i Zuzia poszli do sklepu.
 Tomek.M and Zuzia.F go.PST.3PL.PERS to store
 ‘Tomek and Zuzia went to the store.’

When considering default gender assignment rules, we can look to metalinguistic use of non-nouns, as in (10). Here it appears that the metalinguistic use of *akurat* ‘actually’ triggers neuter agreement on modifiers when used metalinguistically, suggesting neuter gender was assigned to *akurat*.[‡]

- (10) To “akurat” tu nie pasuje.
 DEM.N.SG “akurat” here NEG fit.PRES.3SG
 ‘This “akurat” doesn’t go well here.’

It is worth noting that Polish illustrates why conditions in which a default may be invoked should be determined on a language-by-language basis, as loanwords – typically considered to be a good diagnostic for default gender assignment – in Polish do not take the default (neuter) gender. Instead, loanwords in Polish are typically assigned the masculine animate gender, as shown in (11-b). In fact, [Fuchs \(2014\)](#) argues that modern loanwords from English are categorized as masculine animate, based on their accusative case suffix: masculine animate (but not inanimate) nouns take *-a* in the accusative (11-b). The exception occurs when loanwords end in *-a*, in which case they are femi-

[‡]One might argue that this is due to the fact that *slowo* ‘word’ is neuter, driving the use of neuter when referencing any word metalinguistically.

nine. Interestingly, modern loanwords that end in *-o* are resistant to being classified as neuter, with the plural often being reanalyzed as singular in order to have the *-s* ending that is categorizable as masculine (12).

(11) a. Ten iphone jest wielk-i.
 DEM.M.SG iphone be.PRS.3SG huge-M.SG
 ‘This iphone is huge.’

b. Kupiłam laptop-a.
 buy.PST.ISG.F laptop-ACC.M.ANIM.SG
 ‘I bought (a) laptop.’

(12) Zjadłam po drodze burritos-a.
 eat.PST.ISG.F on road burrito-ACC.M.ANIM.SG
 ‘I ate (a) burrito on the way.’

The assignment of masculine animate to loanwords in Polish is surprising given that membership in this subgender category is thought to be based strictly on semantic features.[§] Fuchs (2014) posits that Polish may have developed an automatic assignment rule, in that loanwords are treated as belonging to a certain gender simply by virtue of being loanwords. Corbett (1991) argues that automatic assignment rules (he addresses the one put forth by Heine (1968)) are undesirable because they are most likely redundant: “Whatever it was that identified [the words] (their morphology or phonology), this information, which is necessary in any case, could equally well provide the means for allocating them to a gender” (Corbett 1991:79). In the case of these English loanwords in Polish, however, morphology and phonology alone would not be sufficient to place a noun in the masculine animate category, for which membership is determined by a combination of phonology and semantics, so there must be some additional mechanism at play. It should be noted that independent evidence for a loanword-specific rule is attested. Boberg (1999, 2009) offers evidence

[§]Fuchs (2014) notes that this is a change in the grammar, as older loanwords like *komputer* ‘computer’ and *telefon* ‘telephone’ are treated as masculine inanimate.

that when loanwords in which the stressed vowel is spelled with <a> are transferred into dialects of North American English, they are realized with the low front vowel /a:/, referred to also as the “foreign (a)”, that the author argues “cannot be explained entirely in phonological terms” ((Boberg 1999:49)). This particular sound has a phonetic realization that is distinct from other sounds in the phonemic inventory of the dialects. The extra-phonemic realization of this vowel in loanwords suggests an automatic assignment rule may be at play, at least at the level at the phonology.

Returning to the more reliable evidence of default gender in Polish, we see that neuter is the default gender in gender assignment. Default gender agreement is also neuter, unless the controller(s) is/are animate, in which case the default may be masculine. This is in line with Kramer (2015)’s prediction that the neuter will be the default gender in all languages with a three-valued gender system, like in Polish. She shows that this is the case for Russian, based on data parallel to what I demonstrated here. This is also a case in which frequency and defaultness do not align – although masculine nouns are more frequent in the Polish lexicon and children learn masculine gender before they learn neuter gender, nevertheless formal evidence points in favor of neuter being the default in gender assignment and gender agreement.

THE NOMINAL PHRASE

The Polish nominal phrase differs from the Spanish phrase in two ways that are critical to this study. First, Polish (like many other Slavic languages) does not have articles – there exists no Polish equivalent of Spanish *el/la*. The second key difference is reflected in many of the examples above. Spanish adjectives are generally postnominal, with a few exceptions, discussed in Section 3.1.1 (13-a). Polish adjectives are largely prenominal (13-b), although they can appear postnominally for stylistic or focus-related reasons (13-c).

(13) a. *Spanish*

el libro rojo ‘the red book’

*el rojo libro

b. *Polish*

czerwona książka ‘(the) red book’

#książka czerwona

Although it is possible for an adjective to occur postnominally in Polish, this is certainly the marked word order within the nominal phrase, requiring special conditions such as contrastive focus, as previously mentioned. In the study detailed in this chapter, the carrier sentences are restricted to the Adj-N word order, for two reasons. The first is of course that the gender information needs to precede the noun in order to be able to test its ability to facilitate lexical retrieval in monolingual and heritage speaker online processing. The second reason is that heritage speakers show a strong preference for rigid word (Song et al. (1997), Montrul (2010b), Polinsky (2008b), O’Grady et al. (2001). Using stimuli with a marked word order would therefore introduce a potential confound, in that heritage speaker processing may be delayed simply because they encounter a scrambled structure.

Taken together, these properties of Polish nominal syntax impact the experimental design in this study. The goal is to determine whether speakers can use gender information to anticipate the lexical item. In Spanish, this gender information was located on immediately prenominal articles *el/la*. In Polish, there are no articles that we can manipulate for this purpose, but there is a wide range of prenominal modifiers that inflect for gender, including demonstratives, numerals, and attributive adjectives, as discussed above. In order to ensure felicitous auditory stimuli, given the task and experimental setup, I determined that color adjectives would be the most natural elements to occur prenominally and host the gender information. The implementation of this choice will be discussed below in Section 4.2.

4.1.2 ACQUISITION OF GENDER IN POLISH

Gender distinctions are broadly considered to be acquired quite early by Polish children: *Smoczynska (1985)* first showed that children have already acquired gender distinctions by age 2;0. This does not necessarily imply that their ability to use nominal inflection paradigms (which require knowledge of gender to be applied correctly) or gender agreement is acquired at the same age, and so subsequent work has refined our understanding of the acquisition of Polish gender.

Krajewski (2005) conducted a corpus analysis of the speech of four monolingual Polish children (ages 1;7-2;6) over the course of 3-4 months. He found that their production of gender markers suggested some of them recognized only three gender categories (masculine, feminine, neuter) while others recognized the subgenres of masculine (animate, inanimate, personal), for a total of five gender categories. Based on these findings, *Krajewski* argues that at the initial stage of gender acquisition, children only have three gender categories (m., f., n.), since these are “global” (every noun must be assigned to one of them), whereas the further distinction between animate, inanimate, and personal nouns applies only within the masculine category.

To determine any asymmetries between the gender categories in acquisition, *Janssen (2016)* compared the acquisition of gender in Polish and Russian children. The Polish children had acquired the distinction between masculine and feminine by age 2;0 but the neuter later on. Still, the Polish children produced more neuter than their Russian counterparts. *Janssen* argues that Polish does not reduce unstressed vowels, so the gender markers on nouns are phonetically transparent, whereas Russian vowel reduction in unstressed syllables makes nominal suffixes often opaque to gender category.

Diminutives and diminutive morphology are thought to play a role in the acquisition of Polish gender, as Polish diminutives maintain the gender of the base noun and are very transparent. *Dabrowska (2006)* investigated the role that diminutives may play in the acquisition of gender by

testing how children performed in using appropriate endings on nonce words that had diminutive endings versus simplex nonce words (ones that could not be analyzed as having an affix). Her experimental results indicate that inflecting the diminutive nonce words was easier for the children, although not by very much. On the other hand, [Janssen \(2014\)](#) conducted an analysis of a corpus of Polish child-directed speech [Haman et al. \(2011\)](#) and found that the frequency of diminutives in child-directed speech is lower than generally thought: 23% of the nouns in the corpus were diminutives.

Moving to bilingual acquisition of gender, [Brehmer & Rothweiler \(2012\)](#) conducted a production study of bilingual Polish-German children (ages 2;11-6;5, split into 6 age groups), eliciting production data of gender-marked adjectives (in contrast to previous studies, most of which focused on inflectional endings on nouns themselves as evidence of appropriate gender assignment). Their participants produced masculine and feminine agreement correctly for both real and nonce words, performing almost at ceiling on real words from about age four. But their results were much lower for the neuter, reaching as low as 0% for neuter nonce words for one of the age groups. Additionally, producing proper gender agreement for masculine and feminine nouns with atypical endings was more difficult than for those with typical endings for the participants, and especially so if there were no semantic cues to the gender of the referent. In making errors, the children mostly overgeneralized the masculine, only occasionally overapplying the feminine gender, which the authors take to be evidence of a default strategy. The two most notable findings of the study are the following: First, since the children performed better on real words than on phonotactically legal and transparent nonce words, the authors argue that children do not solely rely on morphophonological information in determining gender but also on their learned knowledge of the gender of nouns in their lexicon. Second, the error rate even among the oldest bilingual children in this study suggests that the gender system is not completely learned by the time the bilingual child enters the school system. The implications for studies of adult heritage speakers of Polish, if we can extend from this study on

Polish-German bilinguals, is that their gender system is not necessarily fully learned by the time they enter the school system and experience a drastic reduction in the input to the heritage grammar, but that they are able to access the gender information that they have learned without relying solely on morphophonological cues to gender in familiar nouns.

4.2 MATERIALS AND DESIGN

In putting together the auditory and visual stimuli for this study, the goal was to make the design match as closely as possible that of the Spanish study (Chapter 3) and therefore also those of [Lew-Williams & Fernald \(2007, 2010\)](#) and [Grüter et al. \(2012\)](#). The key difference in the designs is rooted in a major syntactic difference between Spanish and Polish, mentioned above: While Spanish has definite determiners (*el/la*), Polish does not. Conversely, Spanish does not have prenominal adjectives, while Polish does. Therefore, in this study, the gender information that participants may or may not use to anticipate the target item was located on a prenominal adjective, as a vowel suffix. It has been shown that monolingual speakers are able to use gender on prenominal modifiers to facilitate lexical retrieval in Russian ([Sekerina \(2015\)](#)), German ([Hopp & Lemmerth \(2018\)](#)), and Dutch ([Loerts et al. \(2013\)](#)), so using adjectives for the Polish study is an appropriate design choice. I decided to use color adjectives, as these were the most felicitous given the nature of the experimental design. This will be discussed further in Section 4.2.2, and possible implications for the interpretation of the results are discussed in Section 5.3.1.

4.2.1 IMAGE SELECTION

The selection of images for the Polish study did not necessitate a norming study. The goals of the Spanish norming study in Section 3.2 were two-fold. First, the norming study needed to control for dialectal variation in noun and/or gender of the candidate images. Spanish is spoken in countries in

Europe, North America, and South America and thus has many dialects scattered around the world. Polish is more or less contained to Poland[¶] and has relatively little variation, or at least little enough that, as a native speaker of Polish, I was able to identify which images would have consistent labels, without expending resources on another norming study.

The second goal of the norming study in Section 3.2 was to ensure that all the images selected for the visual stimuli were consistently identifiable. In other words, that it was clear what each image was supposed to represent. For this, the results of the norming study in Section 3.2 were sufficient to guide the image selection for the Polish study as well, since the property of being identifiable is in most cases not language-dependent.

Again as in the Spanish study, images were selected based on the phonological form of the corresponding nouns. Nouns had to have at least two syllables to allow for looking time, and they had to begin with a consonant, so as to allow for a clear boundary between the gender information on the preceding adjective (which is realized as a suffixal vowel) and the onset of the lexical item itself.

With these considerations and given the 3x3 design of this study, I selected 36 images: 12 masculine, 12 feminine, and 12 neuter. The corresponding words and their frequencies in child-directed speech are listed in Appendix A.2. Because color adjectives are a crucial part of this study (see Section 4.2.3), it was important to balance the colors within the study. For reasons of word-length, three colors were selected for the Polish study: green, red, and blue. Within each gender-set of images, 4 images were blue, 4 were green, and 4 were red. Recall that in the Spanish study, images were one of four colors (red, yellow, green, and blue), but that the color was not a variable of interest, it was merely part of the design of the study intended to anticipate the design of the Polish study, ensuring maximum similarity between the two sets of visual stimuli. It is worth noting that the phonological form of a given gender suffix may differ slightly based on the phonological properties of the adject-

[¶]It is noteworthy, however, that the Polish diaspora in Chicago is so large that Chicago is considered to be the second largest “Polish” city after Warsaw. To my knowledge, there is no study of a Chicago dialect of Polish, but this may be an opportunity for future research.

	M	F	N
red	czerwon-y	czerwon-a	czerwon-e
green	zielon-y	zielon-a	zielon-e
blue	niebiesk-i	niebiesk-a	niebiesk-ie

Table 4.2: In Polish, the phonological form of the gender suffix varies based on the phonological form of the adjective, as presented here for the three adjectives used in the experimental design.

	M	F	N
M	match	mismatch	mismatch
F	mismatch	match	mismatch
N	mismatch	mismatch	match

Table 4.3: The 3x3 design of the visual stimuli in the Polish experiment.

tive itself. This is detailed for the three chosen color adjectives in Figure 4.2, and will be taken into account in the statistical analysis in Section 4.7.

4.2.2 VISUAL STIMULI

There were 108 total visual stimuli. Like in the Spanish experiment, each stimulus consisted of two images (matched for color) with a fixation cross in the center. The design of these stimuli had to take into account the fact that there are three Polish genders, and therefore mismatch conditions had to be balanced such that a target item of one gender appeared in mismatch conditions with each of the other two genders, as schematized in Figure 4.3.

With this in mind, 36 of the visual stimuli were match conditions, and 72 of the visual stimuli were mismatch conditions, balanced such that each gender appeared an equal number of times as the target gender and as the distractor gender. To illustrate, the feminine object *książka* ‘book’ appears as the target item in a match condition (Figure 4.1), a mismatch condition in which the distractor is masculine (Figure 4.2), and a mismatch condition in which the distractor is neuter (Figure 4.3).



Figure 4.1: Polish Match Condition: The target item *książka* 'book' and the distractor *świeczka* 'candle' are both feminine.



Figure 4.2: Polish Mismatch-M Condition: The target item *książka* 'book' is feminine, while the distractor *talerz* 'plate' is masculine.



Figure 4.3: Polish Mismatch-N Condition: The target item *książka* 'book' is feminine, while the distractor *jablko* 'apple' is neuter.

As in the Spanish study, images were paired into visual stimuli not only based on gender but also based on the first phoneme of the lexical item. Images in a visual stimuli had to begin with different consonants, such that the onset of the lexical item was clearly the moment of disambiguation between candidate images in match conditions (as well as in mismatch conditions, if a participant were not attending to the gender information on the color adjective).

4.2.3 AUDITORY STIMULI

Each target item was placed in a carrier phrase schematized in (14). Setting aside the color adjective, the carrier phrase is the Polish translational equivalent of the carrier phrase in the Spanish study, with similar simplicity and light semantic load as in the previous studies (Lew-Williams & Fernald (2007, 2010), Grüter et al. (2012)). All sentences were recorded by a male native speaker of Polish, recently immigrated to the United States from Poland.

- (14) Gdzie jest COLOR-GEN NOUN?
Where is color-GEN noun
'Where is (the) red/green/blue NOUN?'

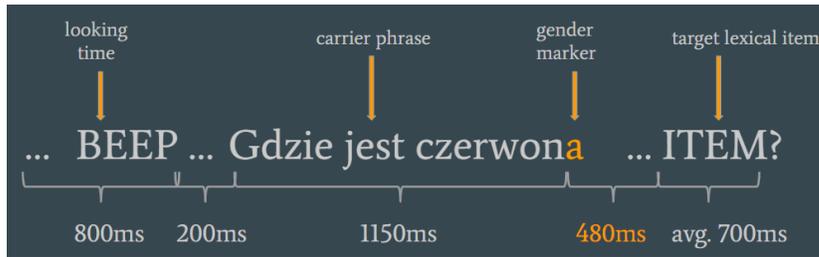


Figure 4.4: Polish auditory stimuli were spliced together to ensure that the onset of the gender information and of the lexical item was at the same time in each stimulus.

The auditory stimuli were constructed by splicing together parts of the recordings, such that the onset of the gender information and the lexical item occurred at the exact same time in every single stimulus, ensuring ease of comparison not only across stimuli within the experiment, but also with the stimuli in the Spanish experiment. The total duration of the looking time plus the beep was the same as in the Spanish study. The carrier phrase in this case consisted of all the spoken information leading up to the gender information, so not just the question words *Gdzie jest...* ‘Where is...’ but also the part of the adjective preceding the gender suffix, as illustrated in Figure 4.4. The duration of this carrier phrase was 1150 ms, as in the Spanish study.

Recall that in the Spanish study, there was 480 ms from the onset of the gender information to the onset of the lexical item. This was true for the Polish study as well, measuring from the onset of the gender suffix on the adjective. The duration of the lexical items was, on average, about 700ms.

Auditory stimuli were paired with visual stimuli, and each stimulus was visible for 6 seconds, with a one-second break between stimuli, during which no images were visible but the fixation cross remained on the screen. Participants were instructed that they could look at the images every time a new set appeared, but that they must direct their gaze to the fixation cross by the end of the beep. This was also the same as in the Spanish study.

4.3 MODEL INTERPRETATIONS

The interpretations discussed here those in Section 3.3 for the Spanish study, adjusted to recognize the presence of three genders in the system. Once again, the dependent variable of interest is the time of first fixation on the target item, indicating the moment in which participants were able to minimize the set of candidate lexical items and thus orient on the target item on the screen. If participants were able to use grammatical gender information of adjectives to anticipate subsequent lexical items, then they should orient on the target item faster in mismatch conditions than in match conditions. However, in this 3x3 design, we need to recognize that not all mismatch conditions were equal: A given target gender could appear in one of two possible mismatch conditions.

(15) Model interpretations:

1: If heritage speakers can use gender information to facilitate lexical retrieval in the presence of a particular distractor gender, then the time of first fixation on target items will be faster in mismatch conditions (in which the distractor is of that given gender) than in match conditions.

2: If heritage speakers cannot use gender information to facilitate lexical retrieval in the presence of a particular distractor gender, then the time of first fixation on target items will be about equal in mismatch conditions (in which the distractor is of that given gender) as in match conditions.

4.4 PARTICIPANTS

Fifty-five Polish speakers aged 18-62 from the Cambridge and Boston area were recruited to participate in the study. For the purposes of this study, the LEAP-Q was translated into Polish.^{||} Partici-

^{||}This document is available in the English version in Appendix B.2 and in the Polish version in Appendix B.4.

pants were asked to fill out the LEAP-Q in either Polish or English. For more information about the LEAP-Q, see Section 3.4.

Based on the number of years they reported living in Poland, participants were identified as belonging to the heritage group or the control group. I used the exact same criteria as in the Spanish study. Those who reported living in Poland for 18 years or more were considered control participants. Those who reported living in Poland for 8 years or less were considered heritage speakers. As in the Spanish study, several participants reported living in Poland for 9-17 years – these are the early L2 learners/speakers (the “1.5” generation) whose onset of bilingualism falls sometime during puberty. A majority of these speakers left Poland to attend boarding school in the UK.

Again as in the Spanish study, a handful of participants were excluded for other reasons. Two participants had neither Polish nor English as their L1. Two participants were excluded for falling outside of the desired age range.

The final sample size was $n = 18$ for the heritage speakers and $n = 23$ for the controls. The age range of these participants is presented in Figure 4.5, and the number of years that the participants reported living in Poland is presented in Figure 4.6.

Table 4.4 provides additional demographic information on the heritage and control participants. As is typically the case for heritage speakers, they spent most of their lives in a Polish-speaking family but very little time in a Polish-speaking work or school environment. It is also notable that the Polish control speakers were highly educated – a product of the fact that many of these speakers responded for to a call for participants circulated by local scientists to their acquaintances.

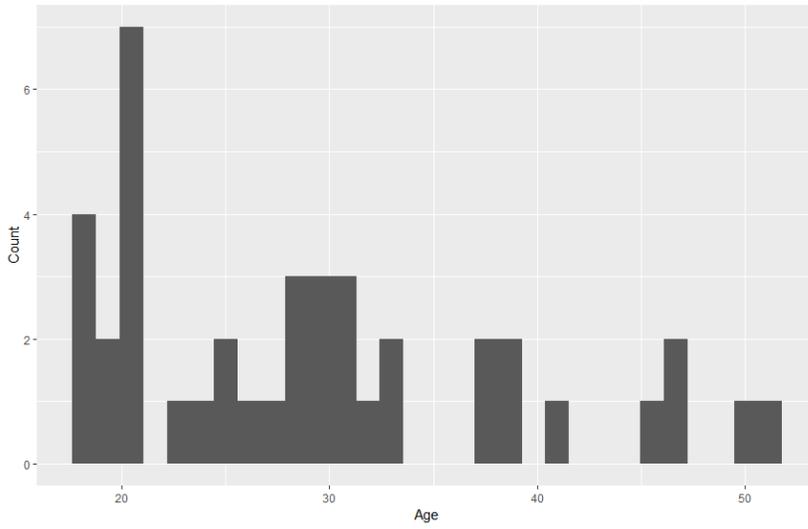


Figure 4.5: A histogram of the ages of Polish-speaking participants in the study whose data was included in the analysis.

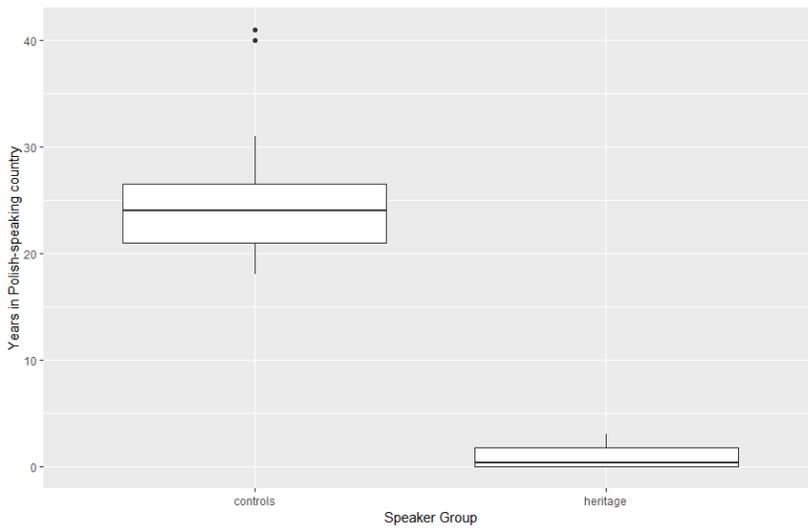


Figure 4.6: The number of years that participants in the heritage group and the control group reported living in a Polish-speaking country.

	<i>n</i>	Age	Years in Spanish-speaking environment			
			Country	Family	Work/School	
Controls	23	31.8 (8.7)	25.0 (8.7)	24.2 (10.8)	19.8 (9.7)	
Heritage	18	26.1 (9.9)	0.8 (1.0)	14.4 (12.5)	1.7 (3.1)	
	Education level					
	HS	Some Coll.	College	Some Grad	Masters	PhD
Controls	1	2	3	0	9	8
Heritage	2	3	8	1	3	1

Table 4.4: Demographic data for Polish-speaking participants, collected from the LEAP-Q. Education level indicates the highest completed degree.

4.5 PROCEDURE

The procedure for this study was the exact same as for the Spanish study. Participants were tested individually in a lab, where they first filled out the LEAP-Q. They then completed a lexical production task, changed from the Spanish study to include only the images used in the Polish study. The participants' responses in the production task were recorded in Praat. Subsequently, participants completed the eye-tracking comprehension task. Participants sat facing a 53.5-cm screen, approximately 75cm away, with their heads placed in an apparatus that ensured minimal head movement during the task. Participants received instructions in written and oral form. They completed four practice trials before the calibration of the SMI EYELINK 1000 Plus; the goal of calibration was set at achieving visual acuity below 0.5 degrees. Gaze position was recorded binocularly at 2000Hz. Participants were allowed to take a break midway through the study, after 54 trials. The length of the break was determined by the participants, and calibration was repeated before the start of the second half of the trials. In total, the study took approximately 45 minutes depending on ease of calibration of the eye-tracker and on how long participants chose to take break. Participants were compensated for their time.

4.6 PROFICIENCY

As in Chapter 3, I used the responses to the lexical production task to approximate the proficiency levels of the control and heritage speakers. Participants viewed a slideshow with the 36 images from the study and were asked to name each image using its color and the noun. The task was self-paced, and the participants were told they could take as much time as they needed to remember the answer, and that they could also skip any image that they could not remember the word for. Participants were also assured that their ability to name the objects would in no way affect their eligibility to take part in the study or the results of the eye-tracking portion. Responses were recorded in Praat.

Correct responses were those for which the noun was an appropriate label for the object in the image (this included diminutives) with the correct gender, and for which the preceding adjective was inflected to match the gender of the noun. For the purposes of determining proficiency, the label that the participants provided did not have to match the noun intended by the experimental design, as long as it was a word that a Polish speaker might use to describe that item.

The proportion of words (out of 36) that each participant labeled correctly is presented in Figure 4.7, split between the controls and the heritage speakers. The control speakers performed effectively at ceiling, with one or two exceptions. Meanwhile, on average, heritage speakers knew 28 words (out of 36), with a standard deviation of 8.8. Not only were the heritage speakers able to recall less of the words overall, they also showed significantly more variation, with some heritage speakers knowing as little as 25% - 50% of the words.

Of the 149 words coded as labeled incorrectly by the heritage participants, 31 were coded as incorrect due to the production of non-target agreement on the color adjective and/or word marker on the noun. Of these, 14 were attributed to a single heritage speaker, who produced masculine agreement on most of the adjectives agreeing with non-masculine nouns. As in Section 3.6, I cannot entirely exclude the possibility that this task is not optimal in terms of sensitivity: it is entirely possible

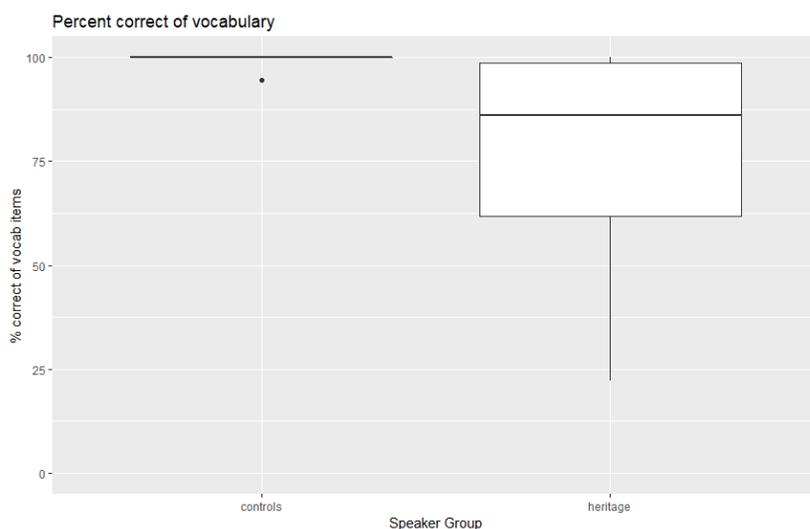


Figure 4.7: The percent of vocabulary items that Polish participants identified using an appropriate label and article.

that some participants' responses may reflect guesses rather than knowledge of gender. Again, given the relative frequencies of the lexical items in child and adult corpora of Polish (Appendix A.2), I can be relatively confident that the lexical items, along with their gender, were familiar to the heritage participants.

As discussed in Section 3.6, the time that it takes participants to perform a word recognition task can also serve as a proxy measure for proficiency. Again, in the case of this task it is an imperfect measure, since some participants waited long periods of time before skipping an item they did not know, whereas others skipped unknown words quite quickly – as a result some heritage speaker completion times are artificially low or artificially high. Nevertheless, we do see some general trends in Figure 4.8. The control group are consistently quick, with a mean time of 100 seconds and a standard deviation of 27.6 seconds. The heritage speakers are unsurprisingly slower overall (mean = 200 seconds) with notably more variation (standard deviation 86.3 seconds).

In addition to the proxy measures of proficiency, the LEAP-Q asks participants to provide self-reported scores on proficiency. Figure 4.9 displays the results for the Polish participants, split across

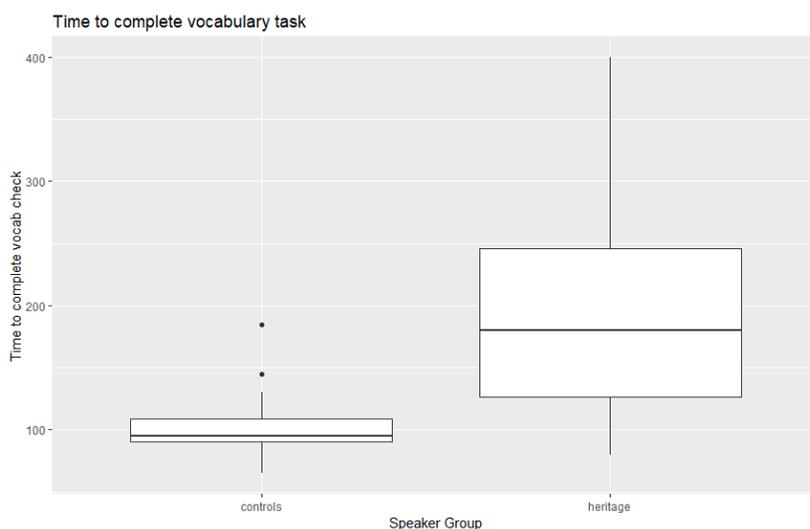


Figure 4.8: The time (in seconds) that it took Polish participants to complete the lexical production task.

control and heritage groups. Participants were asked to rate their proficiency in reading Polish, speaking Polish, and understanding spoken Polish on a scale from 0 to 10 (with 0 being no proficiency and 10 being perfect proficiency). The heritage group was very broadly spread out over the scale, while the controls were clustered at the top end of the scale.

Similarly, participants were asked to rate to what extent they perceive an English accent in their own speech when speaking Polish, and to report how often others identify them as “non-native” speakers of Polish, on a scale from 0 to 10 (0 being never, and 10 being always). Here again, heritage speakers were scattered widely across the scale, while control speakers were clustered at the 0 end of the scale.

One notable aspect of the self-reported proficiency measures is the fact that the control speakers are not consistently rating themselves as fully proficient, and that they’re more likely than the Spanish controls to rate their speech as affected by an English accent. There are two possibilities here. The first is that their proficiency may truly be affected, that their reports are true to their abilities. The second possibility is that this is cultural. In Poland, there is a sense that it is possible to speak

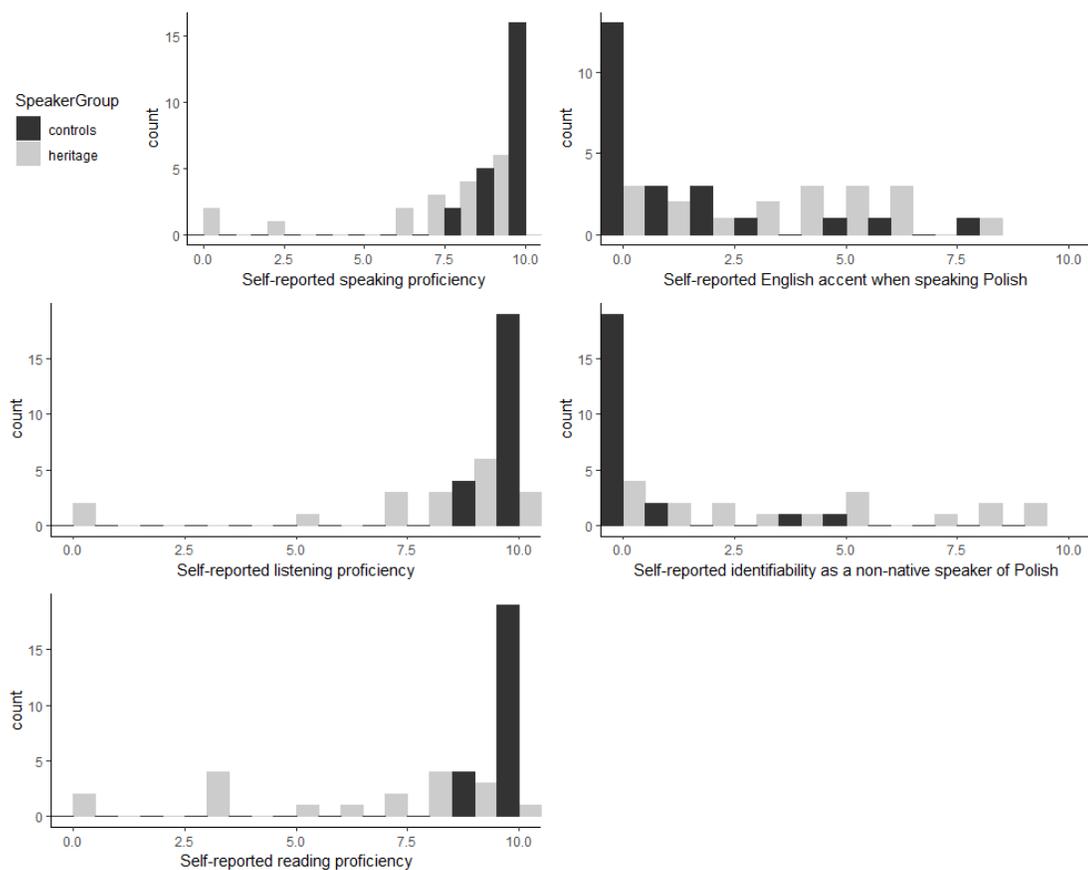


Figure 4.9: Responses to the self-reported proficiency measure questions in the LEAP-Qs. All questions rated on a scale of 0 to 10.

Polish “well” or “beautifully”, or not. That is, even a completely monolingual speaker of Polish may be considered to not speak Polish “well”, if they do not use a particularly sophisticated or poetic syntax and vocabulary. The control speakers’ non-10 ratings here might then be reflective not of their true proficiency but of these cultural standards regarding what it means to be a “good” Polish speaker.

As previously discussed, self-reported proficiency measures are often unreliable for heritage speakers (see Kang & Kim (2012), as well as Chapter 6 of Montrul (2016), and Chapter 3 of Polinsky (2018) for discussion), and due to some cultural norms in Poland, it might be that some of the con-

trols' self-reported proficiency measures may also be unreliable. Therefore, it would be advisable to find a more reliable measure of proficiency, such as the results of the lexical production task. Lexical proficiency has been shown to correlate with syntactic proficiency (Polinsky (1997, 2006), Godson (2003)), and in Section 3.6 I showed that it correlated with the self-reported proficiency measures of the Heritage Spanish speakers. Here I consider how these measures correlate in the Polish data.

Figure 4.10 shows how performance on the lexical production task correlates with heritage speakers' self-reported proficiency measures: reading, speaking, and comprehending spoken Polish, as well as self-perceived English accent in Polish and the frequency of being identified as a “non-native” Polish speaker by others. The speaking, listening, and reading proficiency measures are expected to be positively correlated with performance on the lexical task, given that the upper end of the scales should reflect higher proficiency. Figure 4.10 suggests this may be true, although none of the effects are significant. Significant effects are evident for self-perceived English accent in Polish ($R=-0.5$, $p < 0.05$, Spearman method) and for frequency of being identified as a “non-native” speaker of Polish by others ($R=-0.6$, $p < 0.01$, Spearman method). Since the upper end of the scale for these measures corresponds to lower proficiency, ex. a stronger accent, these correlations are negative.

Figure 4.11 shows the relationship between speed on the lexical production task and the proficiency measures. Again, we see positive but not significant correlations with self-reported proficiency in speaking, listening, and reading. We see significant correlations for English accent in Polish ($R=0.57$, $p < 0.05$, Spearman method) and for frequency of being identified as a “non-native” speaker of Polish by others ($R=0.66$, $p < 0.01$, Spearman method).

The discussion here is similar to the discussion in Section 3.6. The lexical production scores appear to be correlated with the self-reported proficiency measures, but for various reasons the lexical production scores are likely to be more reliable. Therefore, in the analysis in the next section, I will use the results of the lexical production task as a proxy for proficiency in the models in which I test whether proficiency plays a role in how the heritage speakers perform on the eye-tracking task.

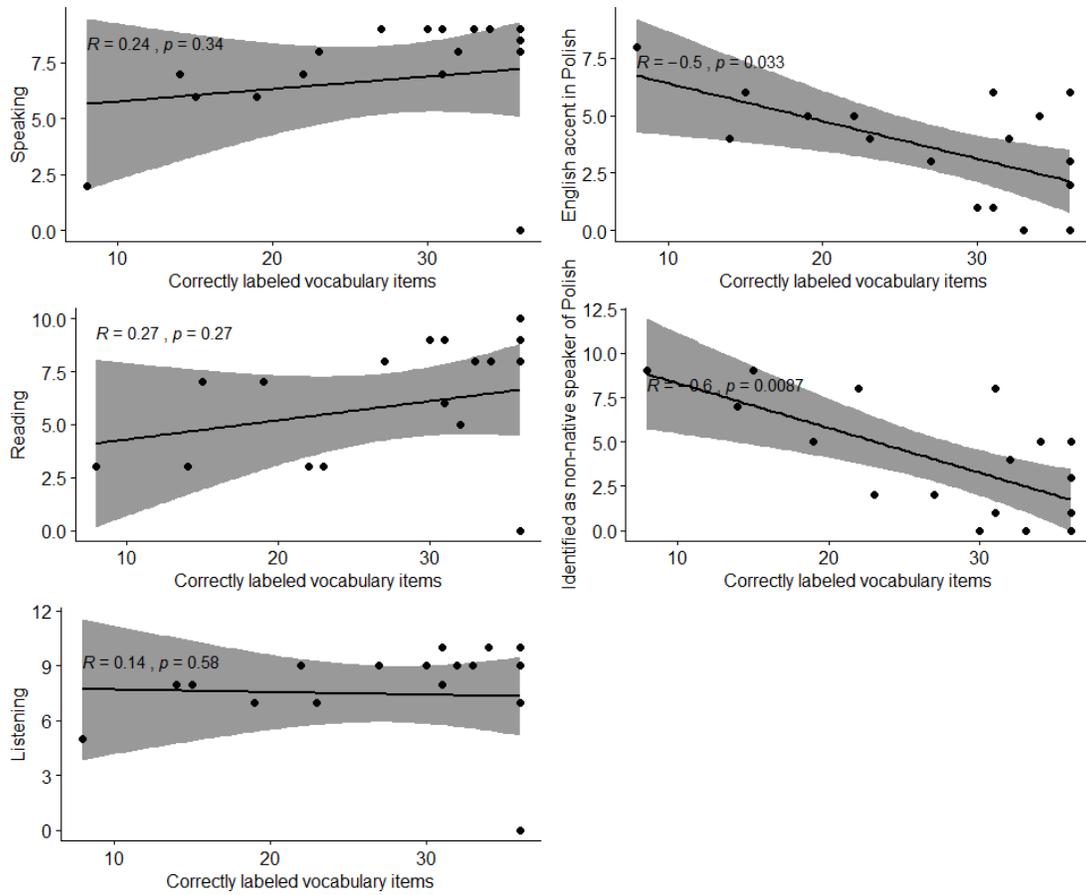


Figure 4.10: Relationship between performance on the lexical production task – in terms of number of vocabulary items (out of 36) that were correctly labeled by the speaker – and the self-reported measures.

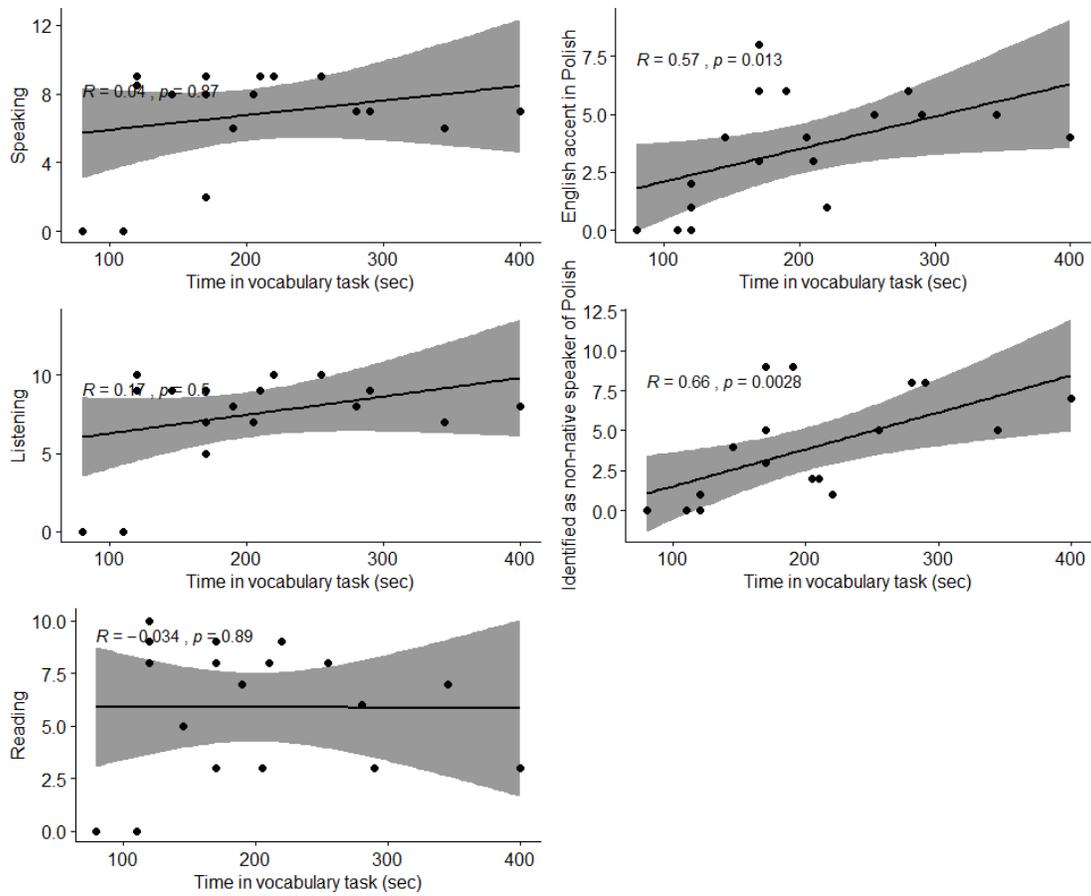


Figure 4.11: Relationship between speed on the lexical production task (in seconds) and the self-reported measures.

4.7 ANALYSIS & RESULTS

Before analyzing the data, for each participants I removed those trials in which the participant did not know the intended lexical item or gender of either (or both) of the images in the visual stimulus. In order to do this, I re-coded the results of the lexical production task to only be “correct” if the participant’s response matched the lexical item intended in the experimental design, and had the appropriate gender. If the participant produced an alternative, it was categorized as incorrect, even if it is an appropriate label in Polish. For instance, some participants identified the picture of the box as *paczka* ‘package’ rather than the intended *pudelko* ‘box’. The only exception to this coding was if the participant provided the diminutive of the intended lexical item. Since Polish diminutives maintain the gender of the base noun, such responses could be considered correct. For instance, several speakers identified the image of the arrow as *strzalka* ‘little arrow’ rather than *strzala* ‘arrow’, but this lexical item has the same gender and (mostly) phonological form, as is the case for most Polish diminutives. This eliminated 140 data points for the control group and 773 data points for the heritage group, as listed in Table 4.5. The higher percentage of eliminated data points in the heritage group makes sense given that they knew a smaller proportion of the lexical items in the oral naming task.

For the remaining trials, data was gathered from the Eyelink software and cleaned and analyzed in R. All fixation times outside of two standard deviations of the mean based on all looking times were trimmed. This left 2248 data points for the controls (having eliminated 93) and 1076 data points for

	Before trimming	After lexical task	After trimming
Controls	2481	2341 (-6%)	2248 (-4%)
Heritage	1918	1145 (-40%)	1076 (-6%)

Table 4.5: The number of data points for both of the speaker groups in the Polish experiment, along with what percent of the data was removed at each stage of data cleaning: removal of unknown or different image labels from the lexical task, and trimming of reaction times outside of two standard deviations of the mean.

the heritage speakers (having eliminated 69). The SR Research Eyelink software records the time of fixation for up to three first fixations on an interest area. For every trial, the earliest fixation in the interest area later than 3250ms after the start of the trial (i.e. the onset of the article) was determined to be the time of first fixation on the target item. The time of first fixation past the critical period of the onset of the gender suffix on the color adjective was compared across match and mismatch conditions and across speaker groups. Since this is a 3x3 design, for ease of analysis and exposition, the results will be presented in three batches, according to the gender of the target item.

4.7.1 FEMININE TARGET ITEM

The mean first fixation times for the control group and heritage group are presented in Figure 4.12. The match condition can be interpreted as in the Spanish study, but the mismatch condition in this study has to be split according to the gender of the distractor: Mismatch-M refers to conditions in which the distractor was masculine, and Mismatch-N refers to conditions in which the distractor was neuter.

I fit a mixed effects linear model predicting time of first fixation by GROUP, CONDITION, and TRIAL, as well as their INTERACTION, with random intercept and slope for CONDITION conditioned on PARTICIPANT and random intercept and slope for CONDITION conditioned on COLOR of the adjective (recall Table 4.6). The categorical CONDITION variable was Helmert contrast-coded to test for a significant difference first between the two mismatch conditions, and then between the mismatch conditions taken together as compared to the match condition.

The model found no significant difference between Mismatch-M (mean = 1080 ms, sd = 347 ms) and Mismatch-N (mean = 1082 ms, sd = 348 ms) ($p = 0.98$), but did find a significant difference between the match condition (mean = 1248 ms, sd = 300 ms) and the two mismatch conditions taken together ($p < 0.01$). Participants' mean time of first fixation was overall faster on mismatch conditions than on the match condition, as illustrated in Figure 4.13. The model also found a significant

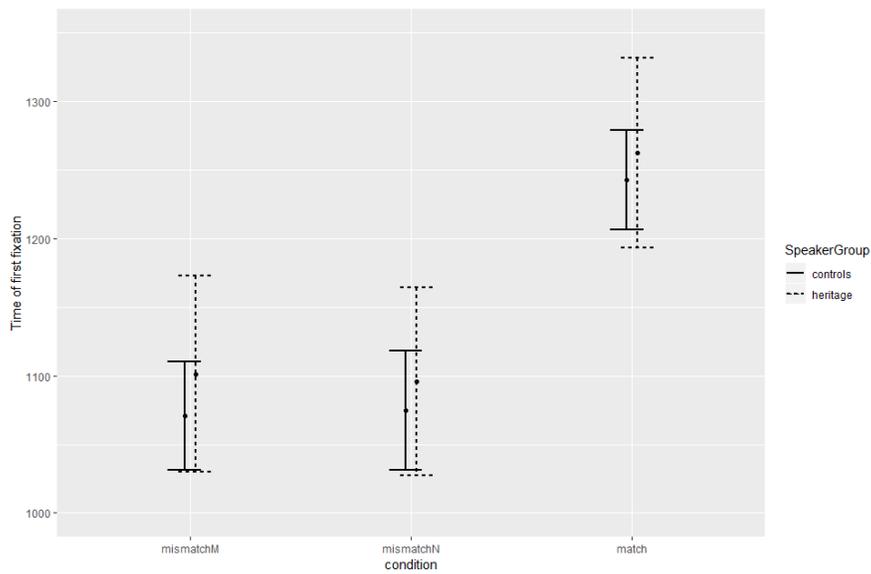


Figure 4.12: Mean time of first fixation on feminine target items (with 95% confidence intervals) across the match condition, and mismatch-M condition, and mismatch-N condition, compared across the control and heritage groups.

effect of TRIAL ($p < 0.01$): the average time of first fixation was faster in later trials. Crucially, the model did not find a significant difference between the heritage group and the control group ($p = 0.41$).

4.7.2 MASCULINE TARGET ITEM

The average time of first fixation for the heritage group and the control group is presented in Figure 4.14. Again, the mismatch conditions are split according to the gender of the distractor – either feminine (Mismatch-F) or neuter (Mismatch-N).

I fit a mixed effects linear model predicting time of first fixation by GROUP, CONDITION, and TRIAL, as well as their INTERACTION, with random intercept and slope for CONDITION conditioned on PARTICIPANT and random intercept and slope for CONDITION conditioned on COLOR of the adjective (Table 4.7). The CONDITION variable was once again Helmert contrast-coded to test for a significant difference between the two mismatch conditions and then between

	<i>Dependent variable:</i>		
	<i>Time of first fixation</i>		
group=heritage	29.476	(35.825)	
condition mismatch M/N	2.035	(13.707)	
condition match	59.278	(11.225)	***
trial number	-46.468	(8.844)	***
group : condition mismatch M/N	3.890	(12.499)	
group : condition match	1.602	(9.710)	
group : trial number	-5.122	(8.842)	
condition mismatch M/N : trial number	14.082	(10.579)	
condition match : trial number	9.956	(6.342)	
group : condition mismatch M/N : trial number	20.567	(10.561)	*
group : condition match : trial number	-2.678	(6.342)	
constant	1,156.308	(35.825)	***
Observations	1,046		
Log Likelihood	-7,290.061		
Akaike Inf. Crit.	14,618.120		
Bayesian Inf. Crit.	14,712.220		

Note:

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table 4.6: Output of the mixed effects model fitted to the Polish data for conditions in which the target item was feminine.

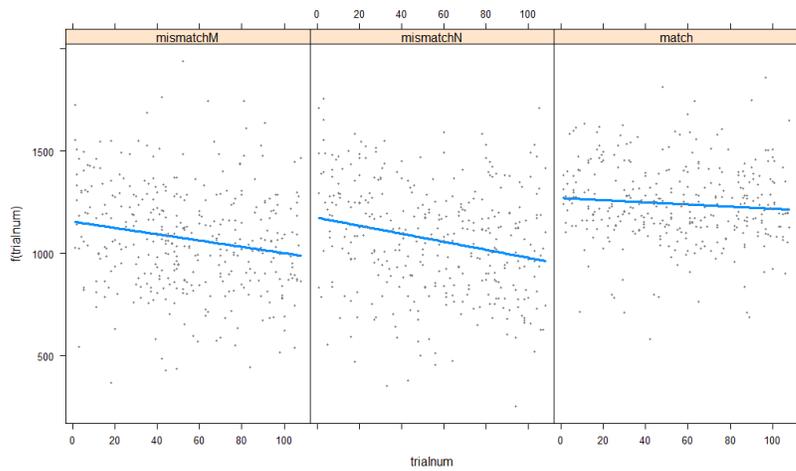


Figure 4.13: The average time of first fixation on feminine target items became faster as the study progressed. There was no significant interaction of condition and trial; therefore, the study cannot conclude that the strength of the effect was different for the three condition types.

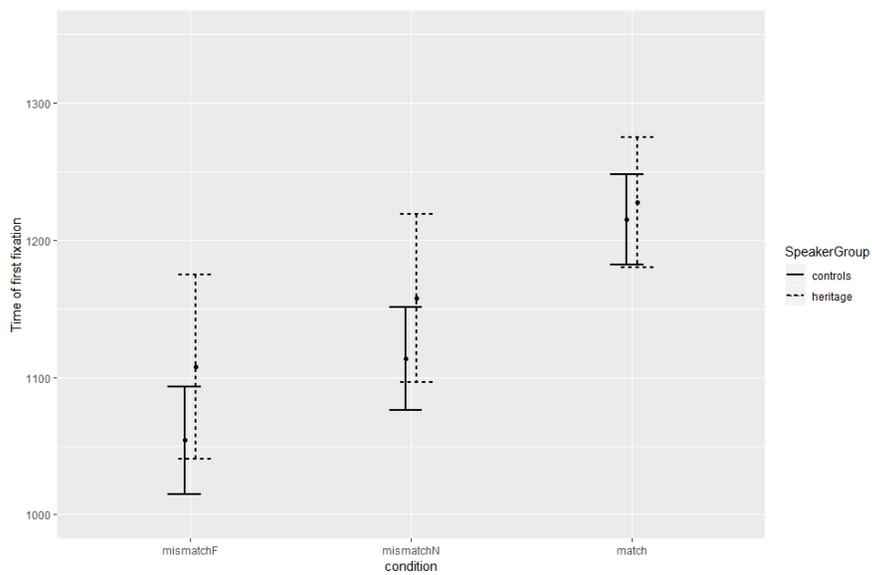


Figure 4.14: Mean time of first fixation on masculine target items (with 95% confidence intervals) across the match condition, and mismatch-F condition, and mismatch-N condition, compared across the control and heritage groups.

	<i>Dependent variable:</i>		
	<i>Time of first fixation</i>		
group=heritage	27.339	(33.126)	
condition mismatch F/N	27.331	(11.929)	**
condition match	38.364	(9.426)	***
trial number	-48.824	(7.139)	***
group : condition mismatch F/N	-2.580	(10.952)	
group : condition match	-6.376	(8.158)	
group : trial number	-6.008	(7.139)	
condition mismatch F/N : trial number	10.440	(8.753)	
condition match : trial number	6.759	(5.080)	
group : condition mismatch F/N : trial number	2.435	(8.726)	
group : condition match : trial number	4.589	(5.078)	
constant	1,155.045	(34.511)	***
Observations	1,162		
Log Likelihood	-7,972.757		
Akaike Inf. Crit.	15,983.510		
Bayesian Inf. Crit.	16,079.610		

Note:

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table 4.7: Output of the mixed effects model fitted to the Polish data for conditions in which the target item was masculine.

the match condition vs the mismatch conditions taken together. The model found a significant difference between Mismatch-F (mean = 1071 ms, sd = 328) and Mismatch-N (mean = 1129 ms, sd = 329 ms) ($p < 0.05$): average time of first fixation on masculine target items was faster when the distractor was feminine than when the distractor was neuter. The model also found a significant difference between the match condition (mean = 1220 ms, sd = 276) and the two mismatch conditions taken together ($p < 0.01$). The model found a significant effect of TRIAL ($p < 0.01$), as illustrated in Figure 4.15. The model found no significant effect of GROUP ($p = 0.41$).

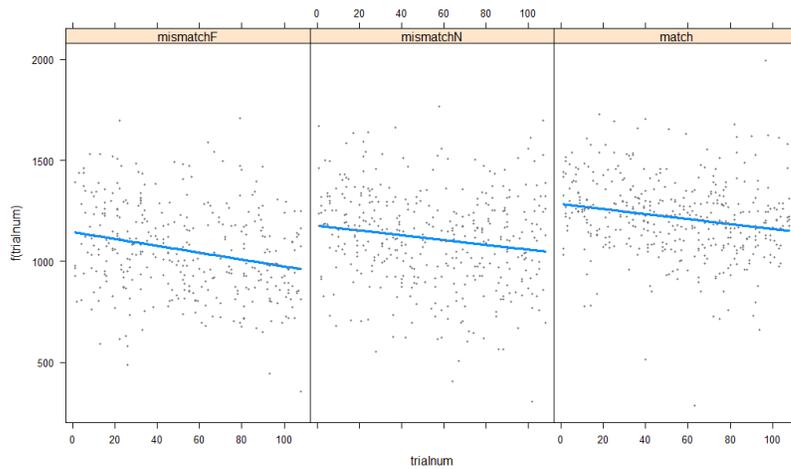


Figure 4.15: The average time of first fixation on masculine target items became faster as the study progressed. There was no significant interaction of condition and trial; therefore, the study cannot conclude that the strength of the effect was different for the three condition types.

4.7.3 NEUTER TARGET ITEM

Figure 4.16 presents the average time of first fixation for the heritage group and control group on match and mismatch conditions when the target item was neuter. Once again, the mismatch conditions are split according to the gender of the distractor item: masculine (Mismatch-M) or feminine (Mismatch-F).

I fit a mixed effects linear model predicting time of first fixation by GROUP, CONDITION, and TRIAL, as well as their INTERACTION, with random intercept and slope for CONDITION conditioned on PARTICIPANT and random intercept and slope for CONDITION conditioned on COLOR of the adjective (Table 4.8). The categorical CONDITION variable was once again Helmert contrast-coded, but this time the coding was slightly different. Exploratory and visual analyses suggest that the Mismatch-M condition behaves more like the Match condition, so the contrast coding was set to first compare Match vs Mismatch-M, and then Mismatch-F vs the other two conditions combined. The model found no significant difference between Mismatch-M (mean = 1166

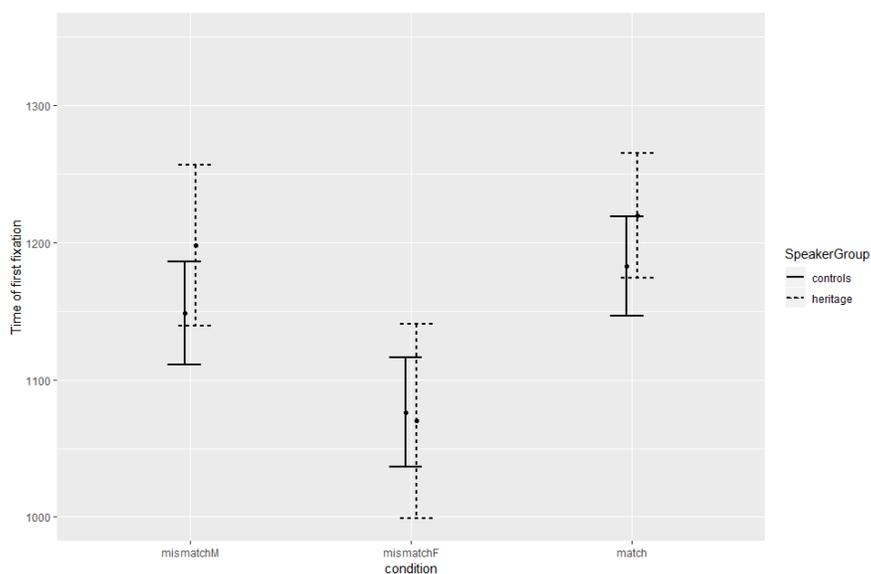


Figure 4.16: Mean time of first fixation on neuter target items (with 95% confidence intervals) across the match condition, and mismatch-M condition, and mismatch-F condition, compared across the control and heritage groups.

ms, $sd = 321$ ms) and Match (mean = 1195 ms, $sd = 279$ ms) ($p = 0.26$), but it did find a significant difference between Mismatch-F (mean = 1074 ms, $sd = 333$ ms) and the other two conditions ($p < 0.01$). The model revealed no significant effect of GROUP ($p = 0.49$) but did show a significant effect of trial ($p < 0.01$), as illustrated in Figure 4.17.

4.7.4 PROFICIENCY MEASURES

As in the Spanish experiment, in order to determine whether proficiency of the heritage speakers predicts their ability to use gender to anticipate the target item, I considered only the subset of the data that came from heritage speakers and fit mixed-effects linear models that included these proficiency measures.

The first set of three models predicted FIXATION by CONDITION, TRIAL, and VOCABSUM, as well as their interaction, for each of the three genders (feminine, masculine, and neuter). The model included a random intercept and slope for CONDITION, conditioned on PARTIC-

	<i>Dependent variable:</i>		
	<i>Time of first fixation</i>		
group=heritage	23.978	(34.746)	
condition mismatch M/match	-15.711	(10.036)	
condition mismatch F	-38.983	(7.514)	***
trial number	-40.159	(7.164)	***
group : condition mismatch M/match	0.306	(10.036)	
group : condition mismatch F	-7.859	(7.514)	
group : trial number	1.679	(7.164)	
condition mismatch M/match : trial number	-7.603	(8.675)	
condition mismatch F : trial number	-4.128	(5.173)	
group : condition mismatch M/match : trial number	-12.403	(8.675)	
group : condition mismatch F : trial number	-2.910	(5.173)	
constant	1,154.564	(34.765)	***
Observations	1,116		
Log Likelihood	-7,640.000		
Akaike Inf. Crit.	15,318.000		
Bayesian Inf. Crit.	15,413.330		

Note:

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table 4.8: Output of the mixed effects model fitted to the Polish data for conditions in which the target item was neuter.

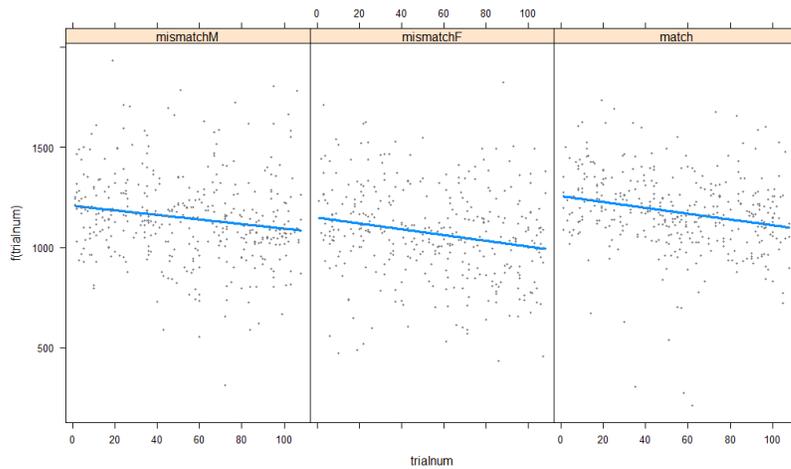


Figure 4.17: The average time of first fixation on neuter target items became faster as the study progressed. There was no significant interaction of condition and trial; therefore, the study cannot conclude that the strength of the effect was different for the three condition types.

IPANT, and a random intercept and slope for CONDITION, conditioned on COLOR of the adjective. The VOCABSUM variable indicated the number of vocabulary items that a participant was able to give an appropriate label for in Spanish. The model did not find a significant effect of VOCABSUM.

The second set of three models predicted FIXATION by CONDITION, TRIAL, and VOCABTIME, as well as their interaction for each of the three genders, and included random intercepts and slopes for CONDITION, conditioned on PARTICIPANT, and a random intercept and slope for CONDITION, conditioned on COLOR of the adjective. Here, VOCABTIME indicates the amount of time participants took to complete the vocab task, in seconds. The models for masculine and neuter target items did not find a significant effect of VOCABTIME. For feminine target items, there was a significant effect of the interaction of CONDITION-MISMATCH and VOCABTIME, as well as the interaction of VOCABTIME, CONDITION, and TRIAL (Table 4.9).

What is interesting in this model is the potential implications for the status of masculine versus

	<i>Dependent variable:</i>		
	<i>Time of first fixation</i>		
condition mismatch M/N	-103.961	(58.881)	*
condition match	109.029	(47.957)	**
vocab time	1.190	(0.922)	
trial number	-23.306	(40.661)	
condition mismatch M/N : vocab time	0.608	(0.309)	**
condition match : vocab time	-0.285	(0.261)	
condition mismatch M/N : trial number	107.089	(50.022)	**
condition match : trial number	-52.799	(28.919)	*
vocab time : trial number	-0.140	(0.231)	
condition mismatch M/N : vocab time : trial number	-0.443	(0.277)	
condition match : vocab time : trial number	0.361	(0.171)	**
constant	963.375	(182.911)	***
Observations	315		
Log Likelihood	-2,177.631		
Akaike Inf. Crit.	4,393.263		
Bayesian Inf. Crit.	4,464.561		

Note: *p<0.1; **p<0.05; ***p<0.01

Table 4.9: Output of the mixed effects model fitted to the Polish data for conditions in which the target item was feminine, with an added fixed effect of the proxy proficiency measure Vocab Time.

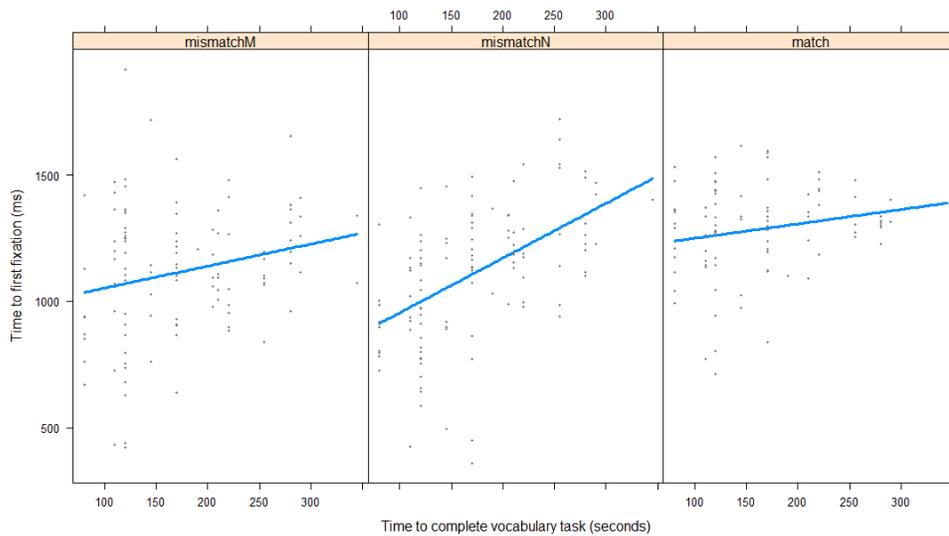


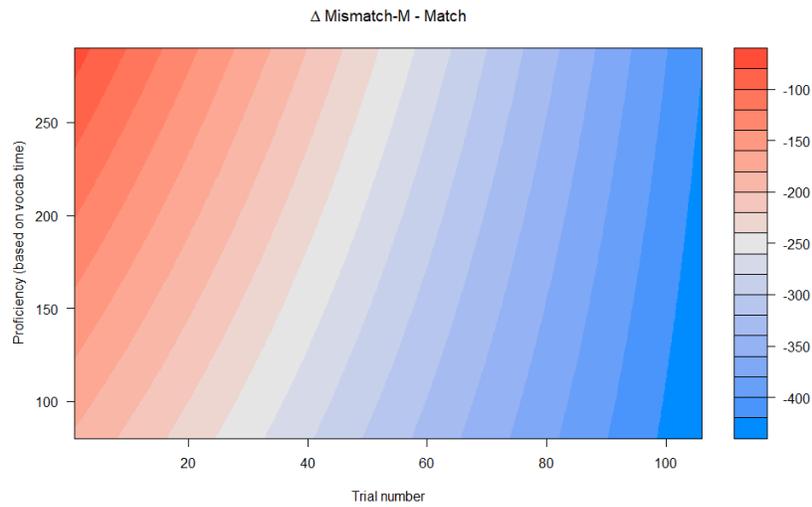
Figure 4.18: The average time of first fixation on feminine target items in three different conditions, based on how long participants took to complete the lexical production task.

neuter. The model for heritage speakers reveals a trend for a main effect of a difference between the two mismatch conditions. It also shows a significant interaction between VOCABTIME and CONDITION-MISMATCH, that is illustrated in Figure 4.18. Generally, heritage speakers who performed slower on the vocabulary task tended to be slower on fixating on feminine target items in mismatch conditions than those who completed the vocabulary task quicker. This effect differs for the two mismatch conditions. It is more pronounced for conditions in which the distractor was neuter: in these conditions, those heritage speakers who performed slower on the lexical task were significantly slower in the eye-tracking task.

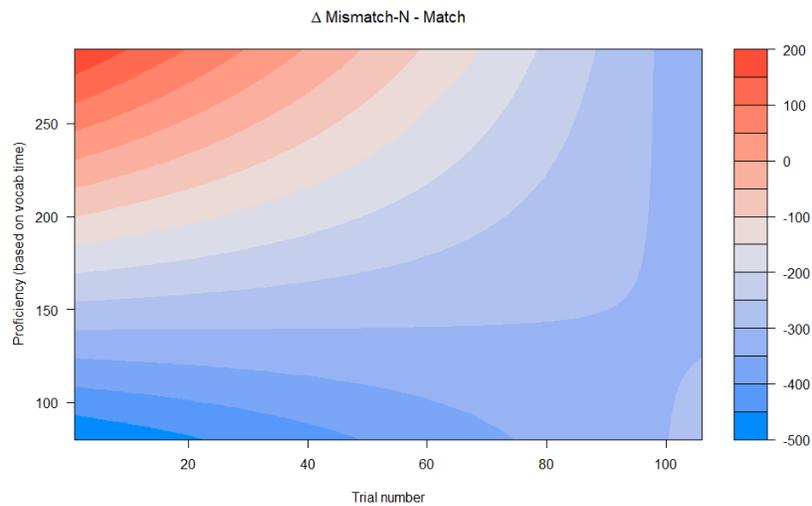
Given the three-way interaction between CONDITION, TRIAL, and VOCABTIME, we should probe this effect further to consider how it changes over the course of the experiment (as trial number increases). To do this, we can investigate how participants' time of first fixation on each mismatch trial compared to that participant's average looking time on feminine match condition trials. These data are represented in the heatmaps in Figure 4.19. In Figure 4.19a we see that the heritage

participants' speed improved at a relatively constant rate over the course of the experiment, regardless of their proficiency level. Proficiency level appeared to dictate their initial speed, but not the rate of change over time. Noting the scale, which indicates all Δ were negative, we see that regardless of proficiency, speakers were able to use the feminine gender in the presence of a masculine distractor. In contrast, Figure 4.19b suggests that the more proficient heritage speakers started out fast and stayed fast, possibly slowing down slightly but still facilitating lexical retrieval (as demonstrated by the negative Δ). On the other hand, less proficiency heritage speakers started out with positive Δ – not being able to use the feminine gender – and rapidly progressed to negative Δ over the course of the experiment, indicating that by the end they were able to use the feminine gender in the presence of a neuter distractor.

We might interpret this to mean that more proficient heritage speakers are able to use feminine gender in the presence of a neuter distractor, but less proficient speakers initially cannot, but can do this later as a learning effect of the study. Meanwhile, regardless of proficiency, the heritage speakers can use the feminine in the presence of a masculine distractor, with a steady learning effect. The implications of these findings for the relationship between feminine and the other genders require further analysis, but they lend tentative support to the contrast between masculine and neuter that will be discussed in the next section.



(a) Difference in looking times between the mismatch condition when the target item is feminine and the distractor is masculine versus the feminine match condition. Note that all Δ are negative, indicating looking times were overall faster on mismatch-M than on match conditions.



(b) Difference in looking times between the mismatch condition when the target item is feminine and the distractor is neuter versus the feminine match condition. Note that the red shades correspond to positive Δ , indicating those looking times for mismatch-N that were slower than looking times for the match condition.

Figure 4.19: Heatmaps illustrating the three-way interaction between proficiency (measured by vocab time), condition, and trial number. Bluer shades indicate faster looking times on the corresponding mismatch condition, while redder shades indicate slower looking times.

4.8 DISCUSSION

The results of the study presented above suggest some asymmetries between the genders – both for the heritage group and the control group. These can be synthesized in the following way:

- When the target noun was feminine, participants were faster in mismatch conditions (regardless of the gender of the distractor) than in match conditions.
- When the target noun was masculine, participants were faster in mismatch conditions than in match conditions. However, between the two mismatch conditions, they were slightly slower when the distractor was neuter than when the distractor was feminine.
- When the target noun was neuter, participants were faster on the mismatch condition when the distractor was feminine than on the match condition. However, they were equally slow on the match condition as on the mismatch condition in which the distractor was masculine.

I will consider in the implications of these asymmetries more closely in Section 4.8.2, but first I will consider the broader research question regarding anticipatory looks among heritage speakers.

4.8.1 FACILITATIVE LOOKS

The goal of this study was to determine whether heritage speakers of Polish can consistently use grammatical gender to anticipate the target noun in the visual world paradigm. Recall the model interpretations:

- (16) 1: If heritage speakers can use gender information to facilitate lexical retrieval in the presence of a particular distractor gender, then the time of first fixation on target items will be faster in mismatch conditions (in which the distractor is of that given gender) than in match conditions.

2: If heritage speakers cannot use gender information to facilitate lexical retrieval in the presence of a particular distractor gender, then the time of first fixation on target items will be about equal in mismatch conditions (in which the distractor is of that given gender) as in match conditions.

We find that (1) holds for all but one gender combination. When the target is neuter and the distractor is masculine, heritage speakers were not able to fixate on the target item faster than in match conditions. In all other mismatch conditions, the gender information on the prenominal modifier allowed the heritage speakers to fixate on the target item faster than in match conditions. Therefore, it is clear that heritage speakers of Polish can access and use grammatical gender on modifiers to facilitate lexical retrieval in online processing.

Notably, the heritage speakers do not differ significantly from the control group in this study. This is true across the board, both in conditions in which they can use gender anticipatorily and in the neuter-target vs masculine-distractor condition in which they cannot. This absence of a slow-down effect is in clear contrast to the slow-down effect seen for heritage speakers of Spanish as compared to the Spanish control group in Chapter 3. This contrast between Polish and Spanish will be the focus of Chapter 6.

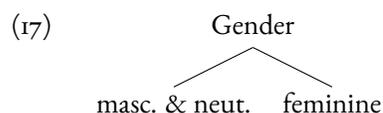
4.8.2 ASYMMETRICAL INTERFERENCE BETWEEN GENDERS

The asymmetries in the results may reveal properties of the gender feature in Polish. As discussed in Section 4.1, the three traditional genders in Polish are masculine, feminine, and neuter. Setting aside the question of subgenders, the three main genders are thought to be in a flat structure, with no hierarchical relationship between them. However, the findings in this study suggest that the representation of the gender feature in speakers' grammars may be more complex than this.

The first distinction that we can make is between the feminine on the one hand and the mascu-

line and neuter on the other. Consider that in any mismatch condition in which the target item was feminine, participants were clearly making anticipatory looks, and in any mismatch condition in which the distractor was feminine, participants were clearly making anticipatory looks. This suggests that a contrast between the feminine and either masculine/neuter is relatively easy for participants in this kind of experimental setting.

On the other hand, a contrast between masculine and neuter is not nearly as easy. Setting aside for a second which gender was the target and which was the distractor, in these contrasts participants were clearly slower. Contrasting this with what we see when one of the items is feminine, we can make an initial split:



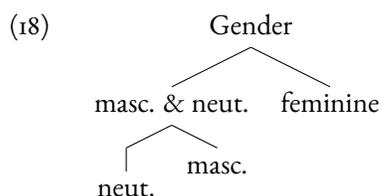
Now we need to probe the relationship between masculine and neuter. The results from the eye-tracking study suggest an asymmetry. When the target is masculine and the distractor is neuter, there is a slight slow-down in anticipatory looks, but they are still anticipatory looks. However, when the target is neuter and the distractor is masculine, the presence of the masculine distractor completely precludes using the neuter gender in an anticipatory manner.

This asymmetrical interference points to the neuter being accessed through the masculine (a more bold proposal might be to say that the neuter is a subtype of the masculine). We can amend the schematization in (17) to include a further distinction, as in (18). Here, the relationship between masculine and neuter is made more specific, indicating a hierarchical relationship between the two. The idea is that when the target is masculine and the distractor is of a subtype or related-but-secondary gender, then this gender only shows partial interference. If however, the target is neuter, and the distractor belongs to the broader, higher-level category, then the presence of this distractor

	masc.	neut.	fem.
nom.	dobry	dobrze	dobrze
gen.	dobrego	dobrego	dobrej
dat.	dobremu	dobremu	dobrej
acc.	dobry	dobrze	dobrze
inst.	dobrym	dobrym	dobrze
loc.	dobrym	dobrym	dobrej

Table 4.10: The declensional paradigm of the Polish adjective *dobry* 'good' in the masculine (inanimate), neuter, and feminine.

attracts enough attention to prevent anticipatory looks to the neuter target.



This is echoed in the morphology of the language. In Table 4.10, we see the declension paradigm of the adjective *dobry* in the three genders: masculine (inanimate), neuter, and feminine. The suffixes for the neuter are syncretic with the masculine in all but the nominative and the accusative (which, moreover, are identical to each other), while the feminine suffixes are completely distinct from both the masculine and the neuter in all cases. Although the nouns in this study appeared only in the nominative, in which masculine and neuter are clearly distinguishable, the results indicate that the syncretism in the remainder of the paradigm may be enough for the two genders to also be linked in the mental representation of the gender values.

An additional consideration for why the feminine stands alone in these results comes from the phonological form of the adjectival inflection. As Table 4.11 shows, in the feminine, the morphological correlate on nouns is the same as the form of the adjectival suffix that agrees in the feminine. This may aid in cuing to the feminine gender feature. In contrast, for the neuter there is only partial over-

	masc.	neut.	fem.
morph. correlate on noun	C	-o/-e/-ę	-a
adjectival suffix (nominative)	-y/-i	-ie/-e	-a

Table 4.11: This table shows, for each gender, the morphological correlates on nouns belonging to that gender class, as well as the shape of the adjectival suffix for adjectives agreeing with nouns of that gender in the nominative.

lap in the phonological form of adjectival suffixes and morphological correlates on the noun, and in the masculine there is no overlap.

The crucial part of this analysis is that it proposes a hierarchical relationship between masculine and neuter, distinct from the feminine. It would be tempting to propose an alternative analysis built on the notion that neuter is the default gender in Polish, per the discussion in Section (4). However, that kind of analysis would make predictions not just about the relationship between masculine and neuter, but also between feminine and neuter. Such an analysis would have to explain why the masculine distractor slows fixation time on neuter targets, but the feminine does not do so.

The analysis presented here echoes [Halle & Matushansky \(2006\)](#), who analyze gender in Russian in a similar way. Russian, like Polish, has a three-gender system, with masculine, feminine, and neuter as the three gender values. Based on adjectival inflection, they provide a featural analysis of the three genders that first groups masculine and neuter as $[-FEM]$ in contrast with feminine $[+FEM]$ before making a further division between masculine and neuter. Polish is a western Slavic language and therefore the Russian analysis cannot be copied wholesale to account for Polish, but the experimental data presented here suggests that the initial split between feminine and non-feminine proposed by [Halle & Matushansky \(2006\)](#) – and reminiscent of the two-gender system in Spanish – is appropriate for Polish as well.

An area for further research is how neuter compares to the other subgenres of masculine. In other words, does the relationship posited here between masculine and neuter extend to other masculine subgenres, or is there a more intricate relationship between them? All of the target items

in this study were inanimate, so the results here bear on the relationship between masculine inanimate and neuter, but one might consider extending this to include the other masculine subgenera. What we know about the acquisition of gender suggests a caveat: the subgenera of masculine are learned relatively late by monolingual children (Krajewski (2005)), so there is a chance that some heritage speakers may not have developed distinctions between some of the subgenera. Nevertheless, from the declensional paradigms we know that the degree of syncretism between the declensional paradigm of the neuter and of the masculine inanimate is mirrored in the other masculine subgenera. Recall from Table 4.1 that the masculine subgenera are distinguished from each other (and therefore also from the neuter) only in the nominative and accusative. This suggests that, at least for monolingual speakers and maybe for some heritage speakers, we may see a similar relationship between neuter and the other masculine subgenera.

Unfortunately, this would be difficult to test in a VWP eye-tracking study. There is some evidence that animate objects or even those implying movement attract the eye more than inanimate/static objects (Howard & Holcombe (2010) and references therein), so this might confound the interpretation of any observed asymmetries between masculine animates and the other two categories. Nevertheless, a more appropriate methodology might help probe the hierarchical relationship between neuter and other masculine subgenera.

4.9 CONCLUSIONS

This study investigated the use of grammatical gender on pronominal modifiers to facilitate lexical retrieval in Polish. To my knowledge, this has not been done for any Polish-speaking population to-date. The results from the Polish control group established that control speakers of Polish are able to use gender to narrow the set of candidate lexical items in anticipation of the subsequent noun, putting them in line with studies that show that monolingual speakers of other European

languages can use grammatical gender on articles in the same way (Lew-Williams & Fernald (2007, 2010), Grüter et al. (2012), Dussias et al. (2013), Hopp & Lemmerth (2018)). Within these overall results, some asymmetries between the genders shed light onto the mental representation of the grammatical gender feature in Polish, suggesting a hierarchical relationship between masculine and neuter.

Results for the heritage speakers show that they can also use gender information on attributive adjectives to anticipate the lexical item, and in fact there is no significant difference between their speed in doing so and that of the control group. While there is no data from L2 speakers for a three-way comparison as there was in the Spanish study, nevertheless the data show quite convincingly that the heritage speakers are monolingual-like in the speed and manner in which they use grammatical gender. This includes the fact that the asymmetrical interference between masculine and neuter appears to be of a similar nature and magnitude for the heritage group as for the control group. The fact that the heritage speakers of Polish show no slowdown as compared to the control group is noteworthy, and the contrast between this finding and the slowdown evident in the results of the heritage Spanish speakers in the previous chapter will be the central focus of the next chapter.

I want to return now to the original motivation for this investigation: the question of whether a more complex gender feature impacts heritage speaker ability to use gender information in on-line processing. More specifically, is a three-gender system (like the one in Polish) more difficult to use than a two-gender system like the one in Spanish. As discussed at the beginning of this chapter, there is reason to think this could go either way. On the one hand, Polinsky (2008c) has shown that heritage speakers of Russian, which has a three-gender system similar to that of Polish, restructure the gender feature to only two categories. Supporting evidence from Polish monolingual and bilingual acquisition suggests Polish children struggle with the neuter, in a way that might be conducive to later restructuring in the heritage grammar. On the other hand, a comparison of gender acquisition of a two-gender system (Dutch) versus a three-gender system (Greek) suggests a more

complex gender system is not necessarily a challenge, and may in fact aid in acquisition due to the system's saliency and transparency (Unsworth et al. (2014)); a similar line of thinking would suggest a three-gender system may be easier to maintain in the heritage grammar.

The data from the Polish study here show that heritage speakers show no significant deviation from the control population in being able to use any of the genders in Polish to facilitate lexical retrieval. This suggests that having a more internally complex gender feature does not make the use of this feature in an online task more difficult. It would be difficult to take this much further and say that the heritage Polish speakers are like the control population (as opposed to the heritage speakers of Spanish, who are slower than their control population) precisely *because* Polish has three genders. As will be discussed in the next chapter, there are many more points of variation between the Spanish study and the Polish study that could be the source of the differences in the heritage/control comparisons. In particular, I would like to focus on the difference between accessing gender on adjectives versus determiners, for which there appears to be independent evidence, although the question remains open as to what might cause the accessing of gender on one to be more or less difficult than the other.

5

Determiners versus modifiers

THERE IS A STRIKING DIFFERENCES between the Spanish study in Chapter 3 and the Polish study in Chapter 4: both groups of heritage speakers can use gender to facilitate lexical retrieval, but the heritage speakers of Spanish do so slower than their control group, while the heritage speakers of Polish are equally fast as the controls. The goal of this chapter is to consider various possible explanations for this contrast. Broadly speaking, there are three possibilities for what might be at the root

of the difference between the Spanish results and the Polish results:

1. Not enough data/evidence to find an effect.
2. Differing levels of proficiency between the heritage speakers of Spanish and the heritage speakers of Polish.
3. The process of accessing gender on determiners (as in the Spanish study) versus on modifiers (as in the Polish study) is inherently different.

Section 5.1 and Section 5.2 will address options (1) and (2) and conclude that while we cannot completely disregard these options, we can consider them unlikely given the available data. Section 5.3 will consider option (3) more thoroughly, but this discussion will ultimately leave open the question of what causes the difference and suggest follow-up studies that target the different hypotheses presented here.

5.1 NULL RESULTS

First we must consider that it is not impossible that heritage speakers of Polish do in fact show a slowdown as compared to the control group, but that this study did not find this difference, due to insufficient data or to a peculiarity of the sample. The lack of a significant difference is referred to as a “null result”, and interpreting null results in frequentist statistics is notoriously tricky. When hypothesis testing yields a low enough p-value, we say that we have a significant difference between populations and can therefore “reject the (null) hypothesis” that there is no significant difference between populations. It is a common misconception that a null result can be interpreted as the lack of a difference between populations. In fact, the nature of frequentist statistics allows us only to interpret a null result as “failing to reject the null hypothesis” that there is no difference between

populations. In other words, a null result implies that we do not have enough reason/evidence to think that there is a difference between the populations.*

In contrast, Bayesian statistics allows us to test the likelihood that a difference between two populations is equal to a specific value. In this case, we would want to test the hypothesis that the difference between the heritage Polish speakers and the control group is equal to zero.† While a Bayesian analysis of the data presented in these studies is beyond the scope of this thesis, we can consider the likelihood that we have an interpretable null effect within the frequentist statistics framework given factors such as our sample size and the size of the effect that we might expect.

An overview of the sample sizes of other work on the anticipatory use of grammatical gender reveals that the effects in this methodology tend to be very robust. Figure 5.1 lists some of the recent work on Spanish (two genders) and German (three genders), along with approximate effect sizes. The purpose of this overview of results is not to compare the effects directly to each other – given the slightly different methodology in each study, that might be inadvisable. Rather, the goal is to consider the sample sizes commonly seen in these studies, along with the magnitude of effects that were found to be significant.

*There are many resources – some specific to language research and some more general – that discuss this very important aspect of hypothesis testing in frequentist statistics. I refer the reader to [Vasishth & Nicenboim \(2016\)](#) for a discussion and further references.

†Again, I point the reader to [Nicenboim & Vasishth \(2016\)](#) for a discussion of the advantages of Bayesian statistics.

Paper	Language	Sample size	Magnitude of (sig.) condition effect
<i>Lew-Williams & Fernald (2007)</i>	Spanish	26 control adults 26 monolingual children	70ms 70ms
<i>Lew-Williams & Fernald (2010)</i>	Spanish	26 control adults 26 L2 adults	90ms not sig.
<i>Grüter et al. (2012)</i>	Spanish	19 control adults 19 L2 adults	60ms not sig.
<i>Dussias et al. (2013)</i>	Spanish	16 control adults 9 L1-English, L2 Spanish 15 L1-Italian, L2 Spanish	120-130ms 90-170ms 200ms
<i>Hopp (2013)</i>	German	20 monolingual adults 9 consistent German L2 11 variable German L2	450 410ms 210ms
<i>Hopp (2016)</i>	German	18 consistent German L2 16 variable German L2	600ms not sig.
<i>Hopp & Lemmerth (2018)</i>	German	15 control adults 24 L1-Russian, L2-German	210-350ms 150ms
[this study]	Spanish	10 control adults 21 heritage adults	200ms 100ms
[this study]	Polish	23 control adults 18 heritage adults	150-200ms 150-200ms

Table 5.1: A sampling of papers on anticipatory use of grammatical gender, with sample sizes and the approximate magnitude of effects across conditions and, where reported, across groups. An overview of these results suggests the effect of interest in these studies is quite robust.

As Figure 5.1 shows, the effects we are looking for are quite robust. Sample sizes range from 9 to 26 participants per test group, and significant effects are consistently found at both ends of the sample-size spectrum. With $n_1=23$ for control Polish speakers and $n_2=18$ for heritage Polish speakers, the sample size for the study that failed to find a significant difference between the two groups in Chapter 4 appears to fit comfortably within the range of sample sizes shown in Figure 5.1 to yield significant effects.

Of course, this is not a guarantee, and there is always a chance that we did not find a difference that does exist between the two populations. Nevertheless, a comparison with other work in this domain shows that the sample size should have been adequate to reveal even a quite small effect, and therefore we can be reasonably confident in moving forward in considering other options for why the Polish heritage speakers do not show a slowdown as compared to the control group, whereas the Spanish heritage speakers do.

5.2 PROFICIENCY

A second factor we must consider is the possibility that the two groups of speakers may have differed in relative level of proficiency. The logic might be that the heritage Spanish speakers might have been less proficient in Spanish than the Polish heritage speakers were proficient in Polish, leading the Spanish group to perform slower than their controls. This is certainly a possibility, given that it is not uncommon for researchers to split heritage speakers within a single study into high- and low-proficiency groups, and find some differences between them.

To address this potential discrepancy, we can compare the relative proficiency measures extracted from the oral picture naming task described in Section 3.6. Figure 5.1 juxtaposes the outcomes for the Spanish control and heritage speakers (left, repeated from Figure 3.7) with the outcomes for the Polish control and heritage speakers (right, repeated from Figure 4.7). Comparing the distribu-

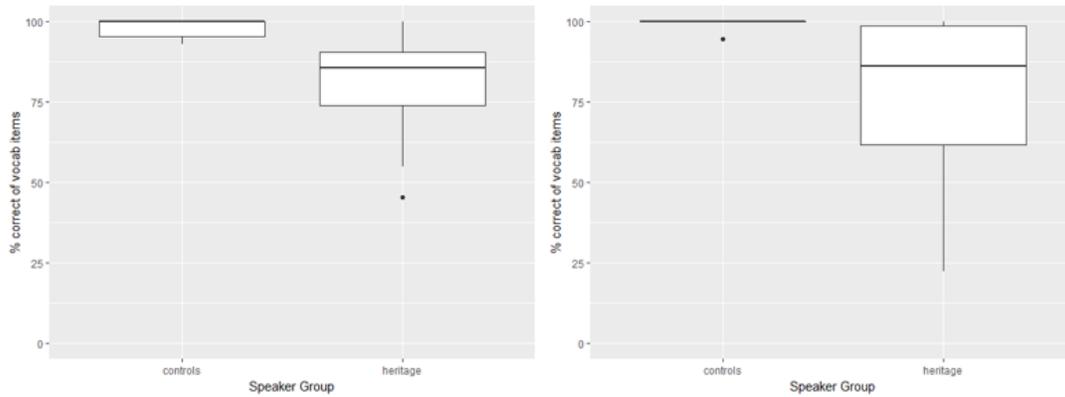


Figure 5.1: The percent of vocabulary items identified correctly by groups in the Spanish study (left) and the Polish study (right). A visual comparison of the two groups of heritage speakers suggests the two groups had comparable proficiency levels overall.

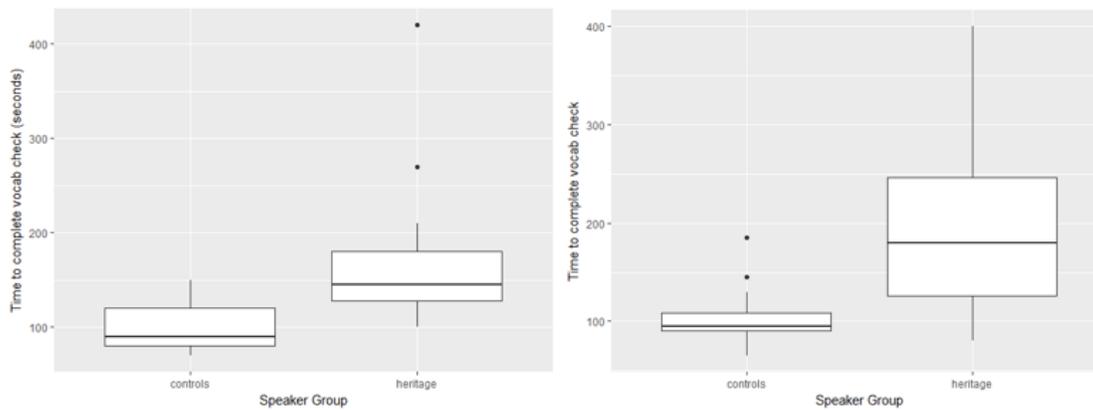


Figure 5.2: The time it took to complete the vocabulary task for speakers of Spanish (left) and Polish (right). A visual comparison of the two groups of heritage speakers suggests the two groups had comparable proficiency levels overall.

tions of the percent of vocabulary items that heritage speakers were able to name correctly, using the appropriate gender, we see that the two groups of heritage speakers display comparable levels of proficiency. If anything, the heritage speakers of Polish show more variation, which is the opposite of what we would expect if proficiency were playing a role.

Similarly, we can compare the time it took these groups to complete the oral picture naming task, as this is also a proxy measure of proficiency. Figure 5.2 allows us to make this comparison between the times for the Spanish-speaking group (left, repeated from Figure 3.8) and the Polish-speaking group (right, repeated from 4.8). To compare the results we have to adjust for the fact that the Spanish task contained 42 images, while the Polish task contained 36 images. Still we see relatively similar distributions for the Spanish and Polish heritage groups. In fact, despite a lower number of images, the Polish heritage speakers appear to have taken more time to complete the task overall, and show greater variation.

As discussed in Section 3.6, both these measures of proficiency have certain drawbacks, but taken together with the fact that the same criteria were used to identify heritage speakers for each sample, they reflect a similar level of proficiency across the two groups. We can therefore likely eliminate the possibility that the lack of a slowdown effect for the heritage speakers of Polish was due to a possible overall higher level of proficiency in the HL as compared to the heritage speakers of Spanish.

5.3 DETERMINERS VS. MODIFIERS

Having addressed the previous two possibilities for why the heritage speakers of Spanish show a slowdown effect as compared to the control group while the heritage speakers of Polish do not, we can consider a final alternative: that it is the difference in the task – accessing gender on adjectives versus definite articles – that is at the root of this difference. Recall that in the Spanish study, the locus of the prenominal gender information is on definite articles *el/la*, whereas in the Polish study,

it is on color adjectives. The possibility I would like to explore here is that accessing the gender information on a definite article is different from accessing gender information on a modifier, and that the former causes difficulty for heritage speakers, whereas the latter does not. Of course ultimately I am comparing two different languages, which is an imperfect strategy. Ideally, the phenomenon would be observed and compared for heritage speakers of a single language. Nevertheless, some considerations offered here will point to a distinction between determiners and modifiers that certainly ought to be pursued further.

This distinction – or rather, the directionality of the distinction – between adjectives and determiners seems counterintuitive, in that it suggests an advantage in accessing adjectives over determiners. Various pieces of evidence would predict the opposite. Definite articles are among the most frequent words in the lexicon, whereas adjectives are far less frequent (more on this in Section 5.3.2). Determiners are largely obligatory – even bare NPs can be argued to have a DP layer – whereas adjectives are optional. We see echoes of this in acquisition: as discussed in Chapter 3, children acquire definite articles early on, and evidence suggests they acquire “proto-determiners” very early (see the discussion in Section 3.1.2). The acquisition of adjectives occurs relatively late in monolingual L1 acquisition (Booth & Waxman (2003)), and agreement on adjectives is acquired after agreement on determiners is mastered (Mariscal (2009)). In heritage languages, adjectives are considered to be prime targets for attrition and loss (Polinsky (2005)).

These broad findings would set the expectation that determiners should be easier to use than adjectives. For additional evidence that there may be a true asymmetry in accessing gender on determiners versus modifiers, we can look to an eye-tracking study that compared adjectives and definite articles within the same language. Unlike in Polish and Spanish, German adjectives and definite articles both appear before the noun. Hopp & Lemmerth (2018) tested L1-Russian, L2-German speakers and monolingual German speakers on the use of grammatical gender on German articles and modifiers to facilitate lexical retrieval of subsequent nouns. Participants in their study encountered

match/mismatch conditions in which pronominal gender information was available on the article only (1-a) and match/mismatch conditions in which this information was available on the adjective only (1-b).

- (1) a. Wo ist der/die/das gelbe NOUN?
where is the.M/the.F/the.N yellow noun
'Where is the yellow NOUN?'
- b. Wo ist ein kleine-r/-s gelbe-r/-s NOUN?
where is a small-M/-N yellow-M/-N NOUN
'Where is a small yellow NOUN?'

While the focus of the paper is on the L2 results, which also show some asymmetries that the authors argue bear on the effect of transfer in second-language learning, the authors briefly note an asymmetry in the monolingual German results as well: “[...] native speakers use gender marking predictively on both articles and adjectives, with the effect being larger for adjectives than for articles” (Hopp & Lemmerth 2018:17). The results offer some evidence that even within a single language, monolingual speakers can access and use gender on modifiers faster than on determiners.[‡]

If this is true for monolingual speakers, then the heritage speaker results would constitute an amplification of something that is already present in a baseline grammar. This is not unusual. Consider, for instance, findings on non-standard use of differential object marking (DOM) in Montrul & Sánchez-Walker (2013). The authors found that the heritage speakers of Spanish living in the US frequently omitted DOM, but that the first-generation immigrants living in the US also produced DOM at a lower frequency than monolingual Spanish speakers living in Mexico. Thus the observations made of the heritage grammar were an amplification of a tendency already present in the control population. A similar example of heritage speakers amplifying a trend that exists in the baseline

[‡]A potential caveat here is that the article condition (1-a) is ambiguous between three genders at the onset of the definite article that carries gender information, whereas in (1-b) *ein* already narrows down the options to just masculine and neuter. The smaller set of possibilities upon encountering gender information in the adjective in (1-b) could play a role in the results.

occurs in Russian case morphology, where Heritage Speakers overgeneralize the genitive plural *-ax* into an oblique plural case marker. This is a reflection of the fact that phonetic changes have made this case marker syncretic across more than one case (Polinsky (2018)). If the findings presented here are accurate, they would fit well with the broader picture of incipient tendencies in monolingual grammars being made more evident in heritage grammars.

The results from German constitute tentative support that the advantage speakers have in accessing gender on adjectives versus determiners exists among monolingual speakers, and it not uncommon for heritage speakers to amplify tendencies already present in the monolingual population. Combining the various pieces of evidence thus points to a difference that ought to be investigated further, as it suggests a few possibilities. The first is that a difference in the frequency of articles and adjectives themselves impacts how they are accessed and processed in an online task (Section 5.3.2), while the second is a difference between functional load based on experience with noun-less constructions in the two languages (Section 5.3.3). A third possibility concerns broader differences between lexical and functional material (Section 5.3.4). While these sections offer some possibilities for follow-ups in Spanish and Polish, it is important to bear in mind that the two languages differ in the syntax of their nominal phrases. Ideally, a study on heritage German or Greek – which have both prenominal determiners and prenominal adjectives – would be useful. At present we do not have enough evidence to adjudicate between the various possibilities discussed in this section, I offer some considerations for further study.

5.3.1 NOTES ON METHODOLOGY

Before turning to possible probabilistic or linguistic explanations of the differences between adjectives and determiners, it is worth acknowledging a methodological distinction, in particular the use of adjectives as the locus of gender information in the Polish study. Adjectives were used to host gender information also in studies on Russian (Sekerina (2015)), German (Hopp & Lemmerth

(2018)), and Dutch (Loerts et al. (2013)), so this design choice is not without precedent in the eye-tracking literature. However, in the study presented here, as in Sekerina (2015)[§], color plays a different role in the visual stimuli than in Loerts et al. (2013) and Hopp & Lemmerth (2018). In the present study (and in Sekerina (2015)), all images in the visual display were of the same color, rendering the color adjective in the auditory stimuli superfluous in terms of providing information that directs the listener to one of the elements in the display. Meanwhile in Loerts et al. (2013) and Hopp & Lemmerth (2018), the color of the images in the display was manipulated such that the color adjective was useful to the participants in narrowing down the set of candidate items as the auditory stimulus unfolded.

One might be concerned that the potential pragmatic oddity of an uninformative adjective in the Polish study may impact how participants engage in the experimental task. Whereas in the Spanish task, the determiners are obligatory, and therefore their presence is entirely expected, adjectives in Polish are optional, and the experimental design of the Polish study makes them pragmatically unnecessary. The Polish participants' attention may have therefore been drawn to the color adjectives, prompting them to think about the significance of this word to the task and discern the goal of the experiment. They may have become aware of the patterns that allow them to use gender to facilitate lexical retrieval, changing their behavior and leading to faster looking times (for a discussion of the impact of awareness on L2 performance in such tasks, see Curcic et al. (2018)).

While I cannot eliminate this possibility entirely, there are a few things to consider. First, the psycholinguistic literature shows that speakers often overspecify, providing their addressees with redundant information regarding certain attributes of a referent. This is found to most likely occur with perceptually salient attributes, in particular color. Koolen et al. (2013) conducted a study in which participants had to explain to another participant which object (target item pre-selected by

[§]Note that the study in Sekerina (2015) has carrier phrases that are declarative statements describing a scene, rather than questions. For the purposes of the Spanish and Polish study, the question carrier phrases replicate previous studies and are a more natural way to ask someone to direct their gaze to an image.

the experimenters) to select out of an array of possible items that varied in size, type, orientation, and color. In a series of tasks, color was never an attribute that would uniquely identify the target item to the listener, and still participants included color adjectives in 10%-30% of utterances. The authors argue that in making decisions about which attributes to use, speakers do not perform a careful analysis of what constitutes a uniquely distinguishing feature but rather rely on other strategies. In particular, the authors suggest that color adjectives are visually salient and therefore immediately grab a speakers' attention, regardless of their contrastive value.

While the results and argumentation in [Koolen et al. \(2013\)](#) do not absolve the Polish study of the methodological concerns regarding the use of color adjectives, they do suggest that redundant color adjectives may be less pragmatically infelicitous in this experimental setting than, say, size adjectives. Following [Koolen et al. \(2013\)](#), color is visually salient and therefore often used in spontaneous production even when it does not assist in ruling out distractors.

A second consideration is that, under the interpretation that the results of the Polish study are a function of the pragmatic infelicity of color adjectives in the carrier phrase, we would have to explain why the outcome is that heritage speakers and controls perform the same. The pragmatic infelicity of the color adjectives should impact the performance of both groups. If the baseline expectation is that the heritage speakers are slower than controls, then a more nuanced story would be required to understand why awareness of the task should level the playing field (as the results show), rather than give both groups a boost (such that the results would look like the Spanish results, but shifted down the time scale).

5.3.2 FREQUENCY OF DEFINITE ARTICLES VERSUS ADJECTIVES

The first possibility mentioned above is that the two processes that are being observed – accessing gender on modifiers and accessing gender on determiners – may be different, leading to the discrepancy. As discussed in Section 3.9, the question has been raised in the literature as to whether

anticipatory use of gender on articles might be a surface phenomenon, rather than a syntactic one (cf. the discussion in [Lew-Williams & Fernald \(2007\)](#)). In other words, it is possible that in these eye-tracking studies in which speakers use the gender information on articles to fixate on the referent earlier in mismatch conditions, they are not truly accessing syntactic gender features but rather relying on probabilistic knowledge, based on the frequency of an article of a particular gender occurring as a unit with a given noun. Under this hypothesis, the observed results ought to be interpreted not as reflecting knowledge and use of gender but rather attention to the probabilistic properties of their language.

By contrast, individual adjectives do not occur with certain nouns at high enough frequencies to posit that participants rely on memorized chunks and probabilities in a VWP task. Instead, with modifiers, participants access the gender feature on the modifier, and this gender feature primes all nouns with a matching gender feature, resulting in faster processing – the “grammatical account” suggested by [Lew-Williams & Fernald \(2007\)](#), following [Friederici & Jacobsen \(1999\)](#) and [Levelt et al. \(1999\)](#).

Interpreting the Heritage Polish and Heritage Spanish results through the lens of this hypothesis, we might say that accessing the syntactic gender feature is equally easy for the heritage speakers and the control group (as seen in the Polish study), but tapping into probabilistic properties of the language slows down heritage speakers, although they are still able to use the probabilities that the control group does (as in the Spanish study). This might admittedly be counterintuitive – *a priori* one might think that accessing surface probabilities should be easier than accessing features in the syntax. However, recall that heritage speakers have less experience with the language overall, and therefore less input on which to build a set of probabilities. The evidence from children’s performance on the study in [Lew-Williams & Fernald \(2007\)](#) lends some credibility to this notion, as children also have less experience with their language (as compared to monolingual adults), and in the study they were slower than adults on the task across the board.

One way of probing the nature of using information on Spanish determiners to facilitate lexical retrieval, we could consider what happens when the gender information on the determiner is at odds with the gender on the noun. In Spanish, feminine nouns that begin with a stressed /a/ must be preceded by the determiner *el*, even though they show feminine agreement otherwise (2-a). This is a strictly phonological effect conditioned on the linear adjacency of the definite article and the noun; the presence of an intervening word allows for the noun to co-occur with the canonical feminine determiner *la* (2-b).

- (2) a. el agua frí-a
 the.M water.F cold-F
 ‘the cold water’
- b. la otr-a agua
 the.F other-F water.F
 ‘the other water’

Testing this class of nouns in the experimental set-up detailed in Chapter 3 would allow us to determine whether the process is probabilistic or whether it entails accessing the grammar. If it is probabilistic, then participants should be able to use the information on *el* to facilitate the retrieval of *agua*, given the frequent co-occurrence of these two elements. If however the process entails accessing the gender feature in the syntax, then participants should struggle: *el* has a masculine gender feature that is in conflict with the feminine feature on *agua*. Pursuing this test would require further consideration, especially since the class of nouns that behaves in this way is limited, and the members of this class show a wide range in frequency of use. Nevertheless, a carefully designed study incorporating this class of nouns could shed light on the processes employed in using gender on articles to facilitate lexical retrieval.

A related possibility is one that draws an inverse relationship between frequency and meaningfulness. Obligatory elements are highly frequent but also not very contentful, whereas adjectives are

less frequent and therefore more meaningful when they do occur. There is independent evidence that the information content of linguistic elements contributes to processing, specifically that more meaningful or more contentful elements are processed faster (Piantadosi et al. (2012, 2018), among others).

Both of these possibilities are rooted in the frequency of individual articles and modifiers. They predict that a nominal element in the same syntactic position as articles *el/la* in Spanish but with much lower frequency in the language ought to close the gap between the heritage speakers of the Spanish and the control group. Such an element would be functionally similar to the definite articles, but its lower frequency would be more akin to that of adjectives. One possibility for such a follow-up study would be to keep the task as it is, but to change the host of the gender feature (and thus also the carrier phrase, to make sure the entire question is felicitous):

- (3) ¿En qué lado hay algún/alguna NOUN?
On which side there.is some.M/some.F NOUN
'On which noun is there some NOUN?'

If the hypothesis detailed here is on the right track, this would make the locus of gender much more similar to that in the Polish study and thus allow the Spanish heritage speakers to perform at similar speeds to the controls. A possible concern with using *algún/alguna* is that the masculine suffix on *algún-∅* is null. Thus, morphologically, the masculine is the absence of a morpheme, the processing of which may be difficult to interpret in an eye-tracking study. One might evade this problem by incorporating plurality into the design, in order to use *algunos/algunas*, for which gender is equally morphologically specified.[¶]

[¶]Note also that *algunos/algunas* are closer in length and semantic content to an adjective than a definite article is, addressing some concerns regarding the head-start speakers potentially have on a multi-syllabic color adjective than on the monosyllabic *el/la* (Christine Shea, p.c.).

5.3.3 FREQUENCY OF NOUNLESS CONSTRUCTIONS

At the root of the difference may also be the utility of determiners in Spanish versus modifiers in Polish in noun-less constructions. In other words, how important or common is it in these two languages to use gender on the nominal element of interest to make inferences about the referent in the absence of a full or overt noun? (For brevity, I'll refer to this (atheoretically) as the functional load of these elements.) Speakers' experience with using gender information in these constructions may impact their performance in the studies presented in the thesis.

In Polish, this question may refer to multiple constructions, in which modifiers appear without an overt noun. The first is commonly referred to as substantivization. In this construction, typically involving contrastive focus, an adjective occurs without an overt noun, but the usual inflection for number, gender, and case is sufficient to identify its referent, as in (4).

- (4) Basia trzym-a czerwon-ą książk-ę, a Kasia (trzym-a) zielon-ą.
Basia hold-PRS.3SG red-ACC.F.SG book-ACC.F.SG CONJ Kasia hold-PRS.3SG green-ACC.F.SG
'Basia is holding the red book, while Kasia is holding the green (one).'

Polish also allows another construction relevant to this discussion: split nominals, illustrated in (5). In these constructions, again typically involving contrastive focus, the modifier may appear disconnected from the nominal phrase, with non-nominal material intervening between the noun and the modifier. Again, the modifier is inflected for concord in number, gender, and case. The same can be said in instances of Left Branch Extraction (cf. [Corver \(2006\)](#) and citations therein) as in (6), in which the adjective precedes the noun it modifies.

- (5) Sukienk-ę mam niebiesk-ą, a spódnic-ę mam
dress-ACC.F.SG have.PRES.1SG blue-ACC.F.SG CONJ skirt-ACC.F.SG have.PRES.1SG
czerwon-ą.
red-ACC.F.SG
'I have a blue dress but a red skirt.'

gender	clitic	adjective
masculine (inan.)	jego/go	dobry
masculine (anim.)	jego/go	dobrego
neuter	je	dobrze
feminine	ją	dobrą

Table 5.2: Direct object forms of Polish pronouns and adjectives.

- (6) Czyj-ą ukrad-łeś książk-ę?
 whose-ACC.F.SG steal-PST,3SG.M book-ACC.F.SG
 ‘Whose book did you steal?’

Split nominals, LBE, and substantivization involve a modifier that occurs without an immediately adjacent overt noun, and so accessing the gender (and number and case) features on the modifier is required to identify the referent in the discourse (in the case of substantivization, this is an overt noun elsewhere in the clause). Thus, accessing the features in this way is a regular task for speakers of Polish, and heritage speakers also likely have experience with these constructions.

Additionally, Polish direct object clitics vary with gender and are used in place of an overt noun, so the gender information on the clitic is used to identify the referent (7). Crucially, the phonological form of these clitics is very similar to that of adjectives, as illustrated in Table 5.2. In the masculine this is only true when the noun is animate, but in the neuter and feminine the animacy split does not apply, and so the direct object clitic is similar in form to adjectives regardless of the animacy of the referent noun.

- (7) Kupi-łeś ją?
 buy-PST,3SG.M CL.DO.ACC.F.SG
 ‘Did you buy it.FEM?’

Articles in Spanish can be used in a similar manner, but not in the exact same constructions. Split nominals do not occur in Spanish, and substantivization requires the presence of a determiner plus at least one adjective. However, the determiners have functional load when used as direct object

of these constructions may be in order. If such a study were to find that split nominals, LBE, substantivization, and direct object clitics are together overall more frequent in Polish than the use of clitics in Spanish, then this might point to a potential explanation of the facts regarding differences in speed of using gender on determiners in Spanish versus modifiers in Polish. To be thorough and take into account the nature of heritage language acquisition, this kind of corpus study would have to balance findings from corpora of adult speech with corpora of adult child-directed speech.

5.3.4 LEXICAL VS FUNCTIONAL MATERIAL

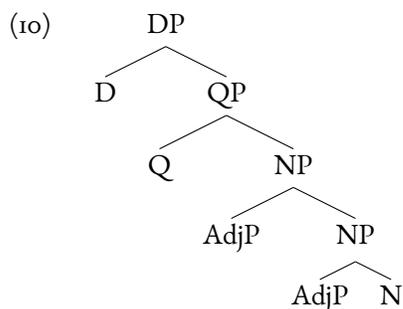
Another candidate for the cause of the difference between determiners and modifiers may be a broader distinction between lexical words and functional words. Broadly speaking, lexical items have semantic content and a meaning of their own, whereas functional items generally express relationships between various parts of a sentence. Lexical and functional items are also distinguished in that lexical items are open-class, indicating that new words may be (and are) added to these categories. Adjectives, nouns, and verbs, are all open-class and readily accept new lexical elements (such as the adjective “woke”, the noun “Facebook”, and the verb “to google”). By contrast, functional elements, such as conjunctions, articles, and prepositions, are closed-class, in that it is rare for there to be a new addition to these categories.

Color adjectives, which hosted the gender feature in the Polish study, are decidedly lexical items. Definite articles, which hosted the gender feature in the Spanish study, are most likely functional elements, as they belong to a closed-class category. Given broad differences between lexical and functional items, it is possible that these differences translate to a contrast in how these elements (or features on these elements) are accessed in online tasks. Evidence from acquisition, mostly described above, points to lexical and function material potentially being learned and stored differently, with determiners in Spanish being initially acquired as chunked units with nouns (and produced as “proto-determiners”) before being parsed as separate morphemes that need to participate

in agreement. No such evidence for chunking or proto-adjectives has been put forth in the literature. It is possible that accessing gender on functional items (in this case, determiners) may be more difficult for heritage speakers than accessing lexical items (adjectives).

It is possible to test this possibility by conducting follow-up studies in both Spanish and Polish, selecting a semantically more contentful but still functional element from each language to be the host of the gender information. Recall the proposal from Section 5.3.2 to use *algún/alguna* in Spanish, as in (9). In Section 5.3.2, where the question under discussion was frequency of elements, these quantifiers were expected to pattern with Polish adjectives under the hypothesis that frequency of determiners versus modifiers is at issue. Here the prediction goes the other way: if the lexical versus functional nature of the host of gender marking is at issue, then Spanish quantifiers (which are functional elements, and located in the syntax as shown in (10)) should pattern with Spanish articles in that heritage speakers should be slower to use gender predictively.

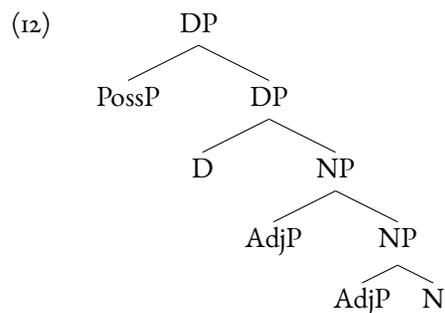
- (9) ¿En qué lado hay algún/alguna NOUN?
 On which side there.is some.M/some.F NOUN
 ‘On which noun is there some NOUN?’



For Polish, the same possibility can be tested by using possessors such as *mój/moja/moje* ‘my,

MASC/FEM/NEUT’.** These are also functional elements, located in the syntax as in (12).^{††} If the lexical/functional distinction outlined here is on the right track, then they should pattern more with Spanish determiners in being harder for heritage speakers to access in the VWP task. And again, if frequency is at issue in the manner discussed in Section 5.3.2, these should pattern with the Polish adjectives in being equally easy for heritage and monolingual speakers to access.

- (11) Gdzie jest mój/moja/moje NOUN?
 where is my.M/my.F/my.N NOUN
 ‘Where is my NOUN?’



There is a potential confound in using the two test cases above to distinguish between the lexical/functional hypothesis and the frequency hypothesis. We refer to feature-matching within the nominal domain as concord to distinguish from a similar – at least at the surface level – phenomenon in the verbal domain, but whether Concord and Agreement are different syntactic processes or not is still a tricky area (cf. discussion in Baker (2008)). If we take seriously the notion that these might be different processes (more on this in a moment), then the lexical/functional distinc-

** As in Spanish, this is not an ideal candidate because phonetic cues to the masculine occur earlier than phonetic cues to the feminine and neuter.

^{††} There are arguments that Polish (and other languages without overt determiners) lack a DP layer (see Bošković (2008, 2009), among others). The status of the NP/DP remains controversial and is outside of the scope of this dissertation. Here, I will include a DP layer in the Polish syntactic trees in order to maintain uniform representations.

tion overlaps with a possible difference in the syntactic processes that give rise of gender features on these elements. So evidence in favor of a lexical/functional distinction would in fact be ambiguous between the lexical/functional difference between determiners versus adjectives and an concord/agreement difference between these two types of elements, respectively.

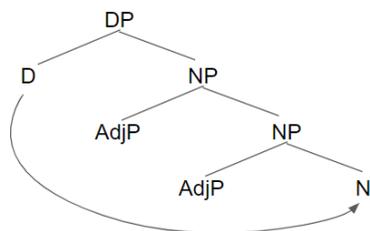
Although much work has conceptualized concord as a formal extension of agreement, [Chomsky \(2001\)](#) and [Chung \(2013\)](#) both allude to the intuition that this need not be the case: concord may be a different syntactic process from agreement. One deeper investigation of this question features in [Norris \(2014\)](#), who outlines the differences between concord and verbal agreement that might motivate such a distinction. The first is that unlike verbal agreement, concord has multiple loci of expression, resulting in feature matching in several elements in the nominal domain. That is, the output of concord can be evident on, for instance, multiple adjectives, whereas the output of agreement is only evident on a single locus (typically the verb). Norris does note that crosslinguistically there are exceptions to this, in that Bantu languages show realization of agreement features on multiple loci in serial verbs (see for illustration [Henderson \(2006\)](#)). Second, the syntactic relationship between the agreeing elements in concord spans heads, specifiers, and adjuncts, whereas verbal agreement is far more structurally restricted – for instance, it is crosslinguistically very rare for adverbs (which are generally assumed to be syntactically similar to adjectives) to enter into agreement. Third, Norris argues that subject-verb agreement is agreement between two different extended projections (a nominal projection and a verbal one), whereas concord is by nature different as it is agreement within a single extended projection. Finally, Norris notes that subject-verb agreement is often thought to be closely tied to case-assignment, whereas no such link is evident for concord.

If we take this distinction seriously, then we might also interpret some findings from the literature on heritage languages as supporting evidence. Recall from the discussion in Section 2.3.4 that heritage speakers show clear differences in their ability to produce and comprehend agreement in the verbal domain versus concord in the nominal domain ([Benmamoun et al. \(2013b\)](#), [Bolonyai \(2007\)](#),

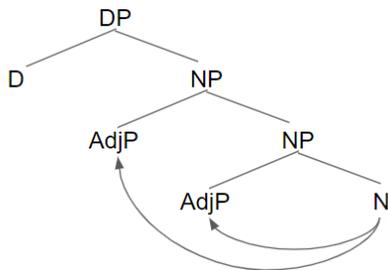
Montrul et al. (2012), Polinsky (2006), Albirini et al. (2011, 2013), Fenyvesi (2000), De Groot (2005)).

Notably, results from this line of research converge on the fact that errors in the nominal domain are more frequent than in the verbal domain, setting the expectation that concord should be more difficult than agreement (much like the evidence from acquisition discussed at the beginning of Section 5.3), contrary to what we have seen in the eye-tracking measures.

If then we follow Chomsky (2001), Chung (2013), and Norris (2014) in considering concord and agreement to be inherently different processes (abstracting away from the precise mechanics of each), then the lexical/functional distinction between adjectives/determiners would be confounded by a possible concord/agreement distinction (respectively). Given the additional assumption that probing for agreement may be restricted to functional material, gender features on determiners would be the result of agreement (13). Meanwhile, adjectives – as lexical material – could only receive gender features via concord (14). Assuming this, the difference in eye-tracking results for determiners and adjectives might be that these systems are processed differently in online tasks.



(13)



(14)

Under the assumptions that concord and agreement need not be two instantiations of the same syntactic process and that probing for agreement is restricted to functional material, the test cases outlined above to test the lexical/functional hypothesis for the difference between adjectives and determiners would not exclude the possibility that the difference is more narrowly the result of two different syntactic processes. First, possessors have also been argued to participate in agreement like determiners (see [Abney \(1987\)](#), [Szabolcsi \(1983\)](#), [Bruening \(2009\)](#), among others). Second, quantifiers – mentioned as a possibility for the Spanish follow-up – behave morphologically like adjectives but may head functional projections. Moreover, while determiners are sometimes thought to participate in agreement rather than concord, there is some evidence that this may be language-specific – the selection of weak vs strong articles in German suggests German determiners participate in concord with other adjectives, and see [Harley \(2013\)](#) for an example in Hiaki.

Now assuming that we can control for these caveats, it might be possible to test for whether this possible contrast between agreement and concord is required. We may expect that predicative adjectives – which show agreement in regular predicates and in small clauses – to behave differently than modifying adjectives that participate in concord in the nominal domain. Pursuing this question could open up an avenue of research with important implications for theoretical analyses of agreement and concord, and for the understanding of possible differences between phi-feature sharing in the v/T/C domain and nominal domain. For the purposes of this dissertation, at present we do not have enough to tease apart the broader lexical/functional distinction versus the more narrow concord/agreement distinction.

5.4 GOING FORWARD

This dissertation leaves open the question as to why heritage speakers of Spanish are slower to use grammatical gender to facilitate lexical retrieval than heritage speakers of Polish. While I cannot

completely rule out the possibility of insufficient statistical power or unmatched proficiency levels between the heritage groups, there is reason to be optimistic that the effect is attributable to some property of the two languages under investigation, specifically the distinction between determiners and modifiers as hosts for gender features within the nominal domain. The exact nature of this distinction is still unclear – is it a frequency effect, a matter of functional load, or an outcome of the syntax? Several possibilities have been outlined in this section, with some potential follow-up studies that may help adjudicate the nature of the difference in the future. In the Conclusion, I interpret this possible distinction within the broader body of evidence pointing to a previously understudied distinction between adjectives and determiners and their participation in agreement within the nominal domain.

6

Conclusion

This dissertation has considered how heritage speakers deploy gender information in online tasks by investigating how heritage speakers of Spanish and heritage speakers of Polish use gender information on pronominal elements to facilitate lexical retrieval in an eye-tracking task in the Visual World Paradigm. While monolingual and bilingual children (at least at an early age) acquire gender with relative ease, adult heritage speakers consistently show non-target-like production and comprehension of gender assignment and agreement, motivating research into what these speakers know about

gender in their HL and how they can use this information in online processing.

Results from the Heritage Spanish speakers demonstrated that they can use gender to facilitate lexical retrieval, in contrast with the L2 speakers studied by [Lew-Williams & Fernald \(2010\)](#) and [Grüter et al. \(2012\)](#). This puts the Heritage Spanish speakers in line with adult monolingual speakers and child monolingual speakers, offering evidence for the notion that a tight link between determiners and nouns develops through early naturalistic acquisition of nominal syntax. This proposal was put forth for L1 vs L2 acquisition based on eye-tracking studies by [Lew-Williams & Fernald \(2010\)](#) and [Grüter et al. \(2012\)](#), and for heritage speakers based on other experimental work by [Montrul et al. \(2014\)](#), but receives novel evidence for heritage speakers here using a much more sensitive measure.

Given that gender is instantiated crosslinguistically with a wide range of gender values, the two-gender system in Spanish served as a starting point, and the three-gender system in Polish was investigated as a natural extension into more complex gender systems. Results from the same study conducted on Heritage Polish showed that these speakers can also use grammatical gender on a prenominal element to facilitate lexical retrieval in a monolingual-like manner. Additionally, asymmetrical interference between masculine and neuter features suggests that the masculine and neuter gender values are in a hierarchical relationship to each other in the mental representation of both the heritage and control Polish grammar. This constitutes novel psycholinguistic evidence to the literature on gender in Polish, which has largely focused on formal investigations of masculine, feminine, neuter, and their subgenres in Polish.

The test of implicit linguistic knowledge and the very sensitive eye-tracking methodology allow us to demonstrate that despite non-target-like production and comprehension of gender assignment and agreement, heritage speakers of Spanish and heritage speakers of Polish can use gender information in a monolingual like manner in online processing. The striking difference between them is also informative. While Heritage Polish speakers use gender to facilitate lexical retrieval just as quickly as

the control group, the Heritage Spanish speakers use gender consistently slower than the controls. This effect may be due to a difference in the process of accessing gender information on determiners versus adjectives.

The exact nature of this process cannot be concluded from the findings in this study, but it does motivate a deeper investigation into the differences between determiners and adjectives in the nominal domain. These differences are certainly echoed in monolingual and bilingual acquisition, in which children's production of gender agreement on determiners is error-free before their production of gender agreement on adjectives. Evidence from code-switching shows a clear preference for code-switching adjectives but maintaining determiners in the HL.

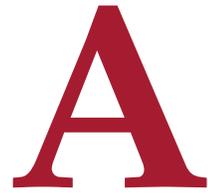
The translation of these facts into performance in eye-tracking and facilitation of lexical access remains an open question. A few possibilities arise, discussed in the previous section, and could be tested using relatively simple extensions of the studies presented here. Further motivation to pursue this line of research is the finding in Hopp & Lemmerth that monolingual speakers of German use grammatical gender on adjectives faster than grammatical gender on articles in a similar task. These results also suggest that similar studies should be conducted on heritage speakers of languages such as German or Greek, in which determiners and modifiers both occur prenominally.

Another outstanding question, mentioned previously in Section 3.8, concerns the role of the dominant language. The heritage speakers tested here all have English as their dominant language; English does not have grammatical gender. It is worth considering whether gender in the dominant language has an impact on heritage speakers' ability to use gender in their heritage language. We can make this question more precise by asking what *about* gender in the dominant language might play a role. It could be that just the presence of a grammatical gender feature in the dominant language might aid heritage speakers in using gender in the heritage language in online tasks, or it could be that what matters is not just presence of gender in the dominant language but also the inner-working of the gender system, ex. the number of feature values.

To test these questions, I replicated the studies here on heritage speakers of Spanish in Germany (Span-Ger) and heritage speakers of Polish growing up in Germany (Pol-Ger). Although I have not yet analyzed the data from these studies, I can consider certain predictions and how I might interpret various possible outcomes of the forthcoming analysis. German has grammatical gender, with three gender categories: masculine, feminine, and neuter. If experience with a grammatical gender feature in the dominant language (regardless of its internal structure) is sufficient to improve heritage speaker ability to access gender in the heritage language, then we may see that the Span-Ger speakers perform faster than those living in the United States. Note that it would be difficult to try to interpret the Polish data in this way, since the Polish heritage speakers living in the US already match the control population in the US. If, on the other hand, the determining factor is not just the presence of gender but also that it categorizes nouns similarly, then I expect the Span-Ger performance to not be impacted compared to the US group, but to find differences between the Pol-Ger group and the US group.

It might be that the differences arise in which genders are susceptible to interference from the gender of the lexical equivalent in the dominant language. [Morales et al. \(2016\)](#) found that bilingual speakers of Italian and Spanish looked less at target items in an eye-tracking study when the gender of the item in the dominant language (Italian) did not match that of the heritage language (Spanish). [Peristeri et al. \(2018\)](#) add data from a production and a comprehension study on Serbian-Greek bilinguals that interference is only evident for late successive bilinguals, but not early successive bilinguals. These outcomes suggest that we may expect the gender of the German equivalent to play a role in heritage speakers' ability to use gender to facilitate lexical retrieval, in that the German gender information is activated in online processing and behaves as a competing cue that slows down lexical retrieval. Whether and to what degree this will be the case for the Span-Ger and Pol-Ger groups remains to be seen, and will shed light on whether different systems of gender categorization impacts the use of gender in online tasks. It should be noted that although [Morales et al.](#)

(2016) did not find an interference effect for monolinguals, there is some evidence that properties of an L2 can impact processing in the L1 (see Dussias et al. (2017) for discussion), and so the studies were also replicated on first-generation immigrants to Germany from Poland and from Spanish-speaking countries.



Experimental stimuli

This appendix contains frequency counts for the stimuli used in the studies presented in this dissertation. Appendix A.1 lists the Spanish stimuli, and Appendix A.2 lists the Polish stimuli. For both lists, two sets of frequencies are provided. The first is from a corpus of adult language, the second is from a corpus of child-directed speech. The reason for including both has to do with the trajectory of heritage language acquisition. The heritage speakers who participated in this study were all adults, but recall from Section 2.3.1 that their lexicon is often restricted in a way that reflects the con-

texts in their lives in which the heritage language was used – often, this is the home, their church, or a community center. Heritage speakers are more likely to be more familiar with vocabulary related to these domains, and so one way to approximate this is to consider the frequency of particular words in child-directed speech, and to compare this to the frequency in adult speech (to illustrate how these frequencies may differ, consider Spanish words *columpio* ‘swing’ or *cadena* ‘chain’). Here, most of the chosen vocabulary items fell within a similar range of frequency.

For the purposes of the analyses conducted in this dissertation, the metric was whether a given heritage speaker knew a given vocabulary item. This was determined through the oral production pre-task in which participants verbally identified each of the images that would later be present in the study, as described in the earlier chapters. For every speaker, any trial for which the speaker did not know one or both of the images was eliminated from analysis, with the result that many of the items listed below that fall in the lower range of relative frequencies were taken out for several of the participants.

A.1 SPANISH STIMULI

This section provides the list of experimental stimuli used in the Spanish study, as described in Section 3.2. The norming study and other criteria used to select these items are provided in Section 3.2 as well. The items are listed here with their corresponding gender in Spanish and with their English translation, along with their frequencies in the Corpus del Español NOW corpus of adult Spanish (5.5 billion segments, 2012 - present) and the CHILDES corpus of Spanish child-directed speech. The CHILDES corpus was narrowed down to include only utterances by a parent (0.7M segments).

	Spanish	Gender	English	CDE-NOW	CHILDES
1.	el casco	masc.	helmet	30	41
2.	el martillo	masc.	hammer	4	29
3.	el gancho	masc.	hook	4	10
4.	el pozo	masc.	well	35	39
5.	el brazo	masc.	arm	62	84
6.	el queso	masc.	cheese	20	412
7.	el sombrero	masc.	hat	11	214
8.	el faro	masc.	lighthouse	9	35
9.	el dado	masc.	dice	2	375
10.	la campana	fem.	bell	10	58
11.	la vela	fem.	candle	17	164
12.	la pala	fem.	shovel	4	47
13.	la bandera	fem.	flag	74	23
14.	la cadena	fem.	chain	129	7
15.	la muñeca	fem.	doll	14	160
16.	la pluma	fem.	feather	10	24
17.	la tetera	fem.	tea kettle	0.2	11
18.	la fresa	fem.	strawberry	4	82
19.	el puño	masc.	fist	9	4
20.	el cerebro	masc.	brain	37	0
21.	el cuchillo	masc.	knife	18	55
22.	el candado	masc.	lock	5	0
23.	el semáforo	masc.	light	13	17
24.	el chaleco	masc.	vest	12	9
25.	el vaso	masc.	cup	16	128
26.	el libro	masc.	book	220	271
27.	el columpio	masc.	swing	1	736
28.	el nido	masc.	nest	7	51
29.	el techo	masc.	roof	43	48
30.	el cuerno	masc.	horn	2	91
31.	la flecha	fem.	arrow	5	10
32.	la rueda	fem.	wheel	98	100
33.	la camisa	fem.	shirt	15	54
34.	la cuchara	fem.	spoon	3	199
35.	la jaula	fem.	cage	6	13
36.	la manzana	fem.	apple	20	237

Table A.1: Stimuli used in the Spanish study, along with their frequency (ppm) in the NOW corpus and in CHILDES (continued on next page).

	Spanish	Gender	English	CDE-NOW	CHILDES
37.	la regla	fem.	ruler	24	9
38.	la casa	fem.	house	661	2484
39.	la bufanda	fem.	scarf	2	44
40.	la carta	fem.	letter	138	61
41.	la mesa	fem.	table	171	880
42.	la cama	fem.	bed	36	1408

Table A.2: (Continued from Table A.1) Stimuli used in the Spanish study, along with their frequency (ppm) in the NOW corpus and in CHILDES.

A.2 POLISH STIMULI

This section provides the list of experimental stimuli used in the Spanish study, as described in Section 4.2. The norming study and other criteria used to select these items are provided in Section 4.2 as well.

The items are listed here with their corresponding gender in Polish and with their English translation, along with their frequencies in parts-per-million (ppm) in the Frequency Dictionary Corpus subpart (0.5M segments) of the corpus of National Corpus of Polish (Narodowy Korpus Języka Polskiego, NKJP) and the Polish Frequency List of Child-Directed Speech (CDS, 0.8M segments) (Haman et al. 2011). The CDS includes only utterances by parents or other caretakers. The numbers here combine frequencies for diminutive and non-diminutive forms of the words, since the corpus considers them to belong to distinct lemmas.

	Polish	Gender	English	NKJP-FDC	CHILDES
1.	koło	neut.	wheel	266	122
2.	pułdo	neut.	box	12	239
3.	mydło	neut.	soap	18	94
4.	wiadro	neut.	bucket	6	101
5.	lustro	neut.	mirror	16	86
6.	drzewo	neut.	tree	120	632
7.	jabłko	neut.	apple	12	430
8.	krzesło	neut.	chair	30	182
9.	łóżko	neut.	bed	60	634
10.	gniazdo	neut.	nest	28	64
11.	pióro	neut.	feather	22	134
12.	jajko	neut.	egg	64	440
13.	rower	masc.	bicycle	14	142
14.	grzebień	masc.	comb	0	31
15.	namiot	masc.	tent	10	10
16.	pasek	masc.	belt	38	103
17.	zegar	masc.	clock	22	193
18.	łańcuch	masc.	chain	18	46
19.	talerz	masc.	plate	24	129
20.	widelec	masc.	fork	2	39
21.	dzbanek	masc.	kettle	4	21
22.	szalik	masc.	scarf	4	27
23.	młotek	masc.	hammer	10	35
24.	samolot	masc.	airplane	118	263
25.	strzała	fem.	arrow	14	57
26.	koszula	fem.	shirt	6	109
27.	linijka	fem.	ruler	4	19
28.	ręka	fem.	arm	200	669
29.	świeczka	fem.	candle	8	94
30.	książka	fem.	book	173	1097
31.	łopata	fem.	shovel	10	54
32.	drabina	fem.	ladder	2	48
33.	huśtawka	fem.	swing	2	55
34.	truskawka	fem.	strawberry	4	49
35.	lalka	fem.	doll	8	595
36.	sukienka	fem.	dress	4	213

Table A.3: Stimuli used in the Polish study, along with their frequencies (ppm) in the NKJP-FDC and CHILDES.

B

LEAP-Q

This appendix contains the Language Experience and Proficiency Questionnaire (LEAP-Q) (Marian, Blumenfeld & Kaushanskaya 2007) as it was presented to participants in the Spanish study in Chapter 3 and the Polish study in Chapter 4. The versions presented here are abbreviated from the publicly available original. The questionnaire contains questions regarding participants' self-reported proficiency and experience with English and either Spanish or Polish. There are also questions pertaining to demographic factors such as education and number of years spent in the US

and outside of the US. Several of these questions provide proxy measures by which to determine whether participants are heritage speakers, control speakers, or members of the 1.5 generation – these questions include number of years in a Spanish/Polish-speaking family and number of years in a Spanish/Polish-speaking school and/or work environment. For reasons associated with task modality, the questionnaire also asked participants to self-report any vision problems, as these may impact performance in an eye-tracking study.

B.1 ENGLISH VERSION FOR SPANISH SPEAKERS

This version of the LEAP-Q is an abbreviated version of the publicly available LEAP-Q. It was adjusted to target Spanish-speakers in the USA and focus on their linguistic experience with English and Spanish.

Northwestern Bilingualism & Psycholinguistics Research Laboratory Marian, Blumenfeld, & Kaushanskaya (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech Language and Hearing Research*, 50 (4), 940-967. Adapted to abbreviated version by Zuzanna Fuchs.

Language Experience and Proficiency Questionnaire (LEAP-Q)

Last name		First name		Today's Date	
Age		Date of Birth		Male <input type="checkbox"/>	Female <input type="checkbox"/>

(1) Please list all the languages you know **in order of dominance**:

1	2	3	4	5
---	---	---	---	---

(2) Please list all the languages you know **in order of acquisition** (your native language first):

1	2	3	4	5
---	---	---	---	---

(3) Please list what percentage of the time you are *currently* and *on average* exposed to each language. *(Your percentages should add up to 100%)*:

List language here:					
List percentage here:					

(4) When choosing to read a text available in all your languages, in what percentage of cases would you choose to read it in each of your languages? Assume that the original was written in another language, which is unknown to you. *(Your percentages should add up to 100%)*:

List language here:					
List percentage here:					

(5) When choosing a language to speak with a person who is equally fluent in all your languages, what percentage of time would you choose to speak each language? Please report percent of total time. *(Your percentages should add up to 100%)*:

List language here:					
List percentage here:					

(6) How many years of formal education do you have? _____

Please check your highest education level (or the approximate US equivalent to a degree obtained in another country):

- | | | |
|--|---|--|
| <input type="checkbox"/> Less than High School | <input type="checkbox"/> Some College | <input type="checkbox"/> Masters |
| <input type="checkbox"/> High School | <input type="checkbox"/> College | <input type="checkbox"/> Ph.D./M.D./J.D. |
| <input type="checkbox"/> Professional Training | <input type="checkbox"/> Some Graduate School | <input type="checkbox"/> Other: |

(7) Date of immigration to the USA, if applicable _____

If you have ever immigrated to another country, please provide name of country and date of immigration here.

(8) Have you ever had a vision problem , hearing impairment , language disability , or learning disability ? (Check all applicable).

If yes, please explain (including any corrections):

Language: SPANISH

This is my (**native** **second** **third** **fourth** **fifth**) language.

(1) Age when you...

<i>began acquiring this language:</i>	<i>became fluent in this language:</i>	<i>began reading in this language:</i>	<i>became fluent reading in this language:</i>

(2) Please list the number of years and months you spent in each language environment:

	Years	Months
A country where this language is spoken		
A family where this language is spoken		
A school and/or working environment where this language is spoken		

(3) Please circle your *level of proficiency* in speaking, understanding, and reading in this language:

Speaking

0	1	2	3	4	5	6	7	8	9	10
None	Very low	Low	Fair	Slightly less than adequate	Adequate	Slightly more than adequate	Good	Very good	Excellent	Perfect

Understanding spoken language

0	1	2	3	4	5	6	7	8	9	10
None	Very low	Low	Fair	Slightly less than adequate	Adequate	Slightly more than adequate	Good	Very good	Excellent	Perfect

Reading

0	1	2	3	4	5	6	7	8	9	10
None	Very low	Low	Fair	Slightly less than adequate	Adequate	Slightly more than adequate	Good	Very good	Excellent	Perfect

(4) In your perception, how much of an English accent do you have in Spanish?

0	1	2	3	4	5	6	7	8	9	10
None	Almost none	Very light	Light	Some	Moderate	Considerable	Heavy	Very heavy	Extremely heavy	Pervasive

(5) Please circle how frequently others identify you as a non-native speaker of Spanish based on your accent in Spanish:

0	1	2	3	4	5	6	7	8	9	10
Never	Almost Never				Half of the time					Always

B.2 ENGLISH VERSION FOR POLISH SPEAKERS

This version of the LEAP-Q is an abbreviated version of the publicly available LEAP-Q. It was adjusted to target Polish-speakers in the USA and focus on their linguistic experience with English and Polish.

Northwestern Bilingualism & Psycholinguistics Research Laboratory Marian, Blumenfeld, & Kaushanskaya (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech Language and Hearing Research*, 50 (4), 940-967. Adapted to abbreviated version by Zuzanna Fuchs.

Language Experience and Proficiency Questionnaire (LEAP-Q)

Last name		First name		Today's Date	
Age		Date of Birth		Male <input type="checkbox"/>	Female <input type="checkbox"/>

(1) Please list all the languages you know **in order of dominance**:

1	2	3	4	5
---	---	---	---	---

(2) Please list all the languages you know **in order of acquisition** (your native language first):

1	2	3	4	5
---	---	---	---	---

(3) Please list what percentage of the time you are *currently* and *on average* exposed to each language. *(Your percentages should add up to 100%)*:

List language here:					
List percentage here:					

(4) When choosing to read a text available in all your languages, in what percentage of cases would you choose to read it in each of your languages? Assume that the original was written in another language, which is unknown to you. *(Your percentages should add up to 100%)*:

List language here:					
List percentage here:					

(5) When choosing a language to speak with a person who is equally fluent in all your languages, what percentage of time would you choose to speak each language? Please report percent of total time. *(Your percentages should add up to 100%)*:

List language here:					
List percentage here:					

(6) How many years of formal education do you have? _____

Please check your highest education level (or the approximate US equivalent to a degree obtained in another country):

- | | | |
|--|---|--|
| <input type="checkbox"/> Less than High School | <input type="checkbox"/> Some College | <input type="checkbox"/> Masters |
| <input type="checkbox"/> High School | <input type="checkbox"/> College | <input type="checkbox"/> Ph.D./M.D./J.D. |
| <input type="checkbox"/> Professional Training | <input type="checkbox"/> Some Graduate School | <input type="checkbox"/> Other: |

(7) Date of immigration to the USA, if applicable _____

If you have ever immigrated to another country, please provide name of country and date of immigration here.

(8) Have you ever had a vision problem hearing impairment language disability or learning disability ? (Check all applicable).

If yes, please explain (including any corrections):

Language: POLISH

This is my (**native second third fourth fifth**) language.

(1) Age when you...

<i>began acquiring this language:</i>	<i>became fluent in this language:</i>	<i>began reading in this language:</i>	<i>became fluent reading in this language:</i>

(2) Please list the number of years and months you spent in each language environment:

	Years	Months
A country where this language is spoken		
A family where this language is spoken		
A school and/or working environment where this language is spoken		

(3) Please circle your *level of proficiency* in speaking, understanding, and reading in this language:

Speaking

0	1	2	3	4	5	6	7	8	9	10
None	Very low	Low	Fair	Slightly less than adequate	Adequate	Slightly more than adequate	Good	Very good	Excellent	Perfect

Understanding spoken language

0	1	2	3	4	5	6	7	8	9	10
None	Very low	Low	Fair	Slightly less than adequate	Adequate	Slightly more than adequate	Good	Very good	Excellent	Perfect

Reading

0	1	2	3	4	5	6	7	8	9	10
None	Very low	Low	Fair	Slightly less than adequate	Adequate	Slightly more than adequate	Good	Very good	Excellent	Perfect

(4) In your perception, how much of an English accent do you have in Polish?

0	1	2	3	4	5	6	7	8	9	10
None	Almost none	Very light	Light	Some	Moderate	Considerable	Heavy	Very heavy	Extremely heavy	Pervasive

(5) Please circle how frequently others identify you as a non-native speaker of Polish based on your accent in Polish:

0	1	2	3	4	5	6	7	8	9	10
Never	Almost Never				Half of the time					Always

B.3 SPANISH VERSION

The Spanish version of the LEAP-Q is an abbreviated version of the publicly available LEAP-Q. It was adjusted to target Spanish-speakers in the USA and focus on their linguistic experience with English and Spanish.

Cuestionario de Experiencia y Competencia Lingüística

Apellido(s)		Nombre(s)		Fecha	
Edad		Fecha de Nacimiento		Masculino <input type="checkbox"/>	Feminino <input type="checkbox"/>

(1) Por favor indique todos los idiomas que conozca **en orden de dominio**:

1	2	3	4	5
---	---	---	---	---

(2) Por favor indique todos los idiomas que conozca **en orden de adquisición** (su idioma materno primero):

1	2	3	4	5
---	---	---	---	---

(3) Por favor indique que porcentaje del tiempo Ud. *actualmente y en promedio* está expuesto a cada idioma.

(Los porcentajes deben de sumar a 100%):

Indique idioma:					
Indique porcentaje:					

(4) ¿Al escoger leer un texto disponible en todos sus idiomas, en que porcentaje de los casos escogería leerlo en cada idioma? Asuma que el texto original fue escrito en un idioma que Ud. no conoce.

(Los porcentajes deben de sumar a 100%):

Indique idioma:					
Indique porcentaje:					

(5) ¿Al escoger que idioma usar para hablar con una persona igualmente fluida a Ud. en todos sus idiomas, que porcentaje del tiempo escogería Ud. hablar en cada idioma? Por favor indique el porcentaje del tiempo total.

(Los porcentajes deben de sumar a 100%):

Indique idioma:					
Indique porcentaje:					

(6) ¿Cuantos años de educación tiene Ud.? _____

Por favor indique su nivel más alto de educación (o la aproximación Estado Unidense equivalente a un título obtenido en otro país):

- | | | |
|--|--|--|
| <input type="checkbox"/> Menos que escuela secundaria | <input type="checkbox"/> Algo de Universidad | <input type="checkbox"/> Maestría |
| <input type="checkbox"/> Escuela secundaria/preparatoria | <input type="checkbox"/> Universidad | <input type="checkbox"/> Ph.D./M.D./J.D. |
| <input type="checkbox"/> Entrenamiento Profesional | <input type="checkbox"/> Algo de Escuela Post-Graduado | <input type="checkbox"/> Otro: |

(7) Fecha de inmigración a los Estados Unidos, si aplicable _____

Fecha de inmigración a los Estados Unidos, si aplicable _____

(8) ¿Ud. ha tenido un problema de visión , impedimento de audición , incapacidad de lenguaje , o incapacidad de aprendizaje ? (Indique todo lo aplicable).

Si es el caso, por favor explique (incluyendo cualquier corrección/es necesaria/s):

Idioma: ESPAÑOL

Este es mi idioma (**materno** **segundo** **tercero** **cuarto** **quinto**).

(1) Edad cuándo Ud. ...

<i>empezó a adquirir:</i>	<i>llegó a ser fluido en:</i>	<i>empezó a leer en:</i>	<i>llegó a leer fluidamente en:</i>

(2) Por favor indique el número de años y meses que Ud. pasó en cada ambiente lingüístico:

	Años	Meses
Un país donde este idioma es hablado		
Una familia donde este idioma es hablado		
Una escuela y/o ambiente de trabajo donde este idioma es hablado		

(3) Por favor seleccione su *nivel de competencia* al hablar, comprender, y leer este idioma:

Hablar

0	1	2	3	4	5	6	7	8	9	10
Ninguna	Muy baja	Baja	Pasable que adecuada	Poco menos adecuada	Adecuada	Poco más adecuada	Buena	Muy buena	Excelente	Perfecta

Comprender

0	1	2	3	4	5	6	7	8	9	10
Ninguna	Muy baja	Baja	Pasable que adecuada	Poco menos adecuada	Adecuada	Poco más adecuada	Buena	Muy buena	Excelente	Perfecta

Leer

0	1	2	3	4	5	6	7	8	9	10
Ninguna	Muy baja	Baja	Pasable que adecuada	Poco menos adecuada	Adecuada	Poco más adecuada	Buena	Muy buena	Excelente	Perfecta

(4) ¿Según a su percepción, cuánto acento extranjero tiene Ud. en español?

0	1	2	3	4	5	6	7	8	9	10
Ninguno	Casi ninguno	Muy ligero	Ligero	Algo	Moderado	Considerable	Pesado	Muy pesado	Extremamente pesado	Penetrante

(5) Por favor valore que tan frecuentemente los demás lo identifican a Ud. como un hablante no nativo de español basado en su acento en español:

0	1	2	3	4	5	6	7	8	9	10
Nunca	Casi nunca			La mitad del tiempo						Siempre

B.4 POLISH VERSION

This version of the LEAP-Q was translated from the English abbreviated version in B.2 into Polish by Ewa Brzezińska, who was compensated for her time. It was adjusted to target Polish-speakers in the USA and focus on their linguistic experience with English and Polish.

Northwestern Bilingualism & Psycholinguistics Research Laboratory Marian, Blumenfeld, & Kaushanskaya (2007).
 The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. [Kwestionariusz Doświadczenia i Biegłości Językowej (LEAP-Q): Ocena profilu językowego u osób dwu- i wielojęzycznych.] *Journal of Speech Language and Hearing Research*, 50 (4), 940-967. Przygotowany w wersji skróconej przez Zuzannę Fuchs. Przetłumaczony przez Ewę Brzezińską.

Kwestionariusz Doświadczenia i Biegłości Językowej (LEAP-Q)

Nazwisko		Imię		Data	
Wiek		Data urodzenia		Mężczyzna <input type="checkbox"/>	Kobieta <input type="checkbox"/>

(1) Proszę wymienić wszystkie języki, jakimi się posługujesz, według **poziomu ich znajomości**:

1	2	3	4	5
---	---	---	---	---

(2) Proszę wymienić wszystkie języki, jakimi się posługujesz, w **kolejności nauczania się** (jako pierwszy język ojczysty):

1	2	3	4	5
---	---	---	---	---

(3) Proszę określić w jakim stopniu masz *obecnie przeciętnie* kontakt z każdym z tych języków (*wartości procentowe powinny sumować się do 100%*):

Wymień języki:					
Określ wartość %:					

(4) W przypadku możliwości wyboru przeczytania tekstu w każdym z tych języków, w jakim odsetku przypadków wybierzesz przeczytanie go w tym konkretnie języku? Załóżmy, że wersja oryginalna powstała w innym, nieznanym dla Ciebie języku (*wartości procentowe powinny sumować się do 100%*):

Wymień języki:					
Określ wartość %:					

(5) W przypadku możliwości rozmowy z osobą, która mówi w tym samym stopniu płynnie w każdym ze znanych Ci języków, w jakim odsetku sytuacji wybierzesz rozmowę w tym konkretnie języku? Proszę wskazać odsetki wszystkich sytuacji. (*wartości procentowe powinny sumować się do 100%*):

Wymień języki:					
Określ wartość %:					

(6) Ile lat kształcenia masz za sobą? _____

Proszę wskazać stopień osiągniętego wykształcenia (lub wskazać amerykański odpowiednik wykształcenia zdobytego w kraju innym, niż USA):

- | | | |
|---|---|---|
| <input type="checkbox"/> Podstawowe (poniżej High School) | <input type="checkbox"/> Rozpoczęta szkoła wyższa/College | <input type="checkbox"/> Wyższe magisterskie |
| <input type="checkbox"/> Średnie (High School) | <input type="checkbox"/> Licencjat (ukończony College) | <input type="checkbox"/> Doktorat (Ph.D./M.D./J.D.) |
| <input type="checkbox"/> Zawodowe | <input type="checkbox"/> Rozpoczęte studia podyplomowe | <input type="checkbox"/> inne: |

(7) Data imigracji do USA – jeśli dotyczy _____

Jeśli dotyczy Cię imigracja do innego kraju, wskaż proszę jego nazwę oraz datę imigracji poniżej:

(8) Czy masz zaburzenia wzroku □, słuchu □, zaburzenia językowe □ lub zaburzenia uczenia się □? (Zaznacz wszystkie pasujące).

Jeśli tak, opisz je (w tym próby

korygowania): _____

Język: POLSKI

To (pierwszy (rodzimy) drugi trzeci czwarty piąty) nauczony przeze mnie język.

(1) Wiek kiedy ...

Rozpocząłeś/rozpoczęłaś naukę tego języka:	Osiągnąłeś/osiągnęłaś płynność językową:	Rozpocząłeś/rozpoczęłaś czytanie w tym języku:	Osiągnąłeś/osiągnęłaś płynność w czytaniu w tym języku:

(2) Proszę określić liczbę miesięcy i lat spędzonych w podanym środowisku językowym:

	Lat	Miesiące
Kraj w którym mówi się w tym języku		
Rodzina w której używa się tego języka		
Szkoła i/lub praca, w której używa się tego języka		

(3) Proszę zakreśli *poziom biegłości* w tym języku w zakresie mówienia, rozumienia oraz czytania:

Mówienie

0	1	2	3	4	5	6	7	8	9	10
Zerowa	Bardzo niska	Niska	Niezła	Odrobinę poniżej zadowalającej	Zadowalająca	Odrobinę powyżej zadowalającej	Dobra	Bardzo dobra	Świetna	Doskonała

Rozumienie języka mówionego

0	1	2	3	4	5	6	7	8	9	10
Zerowa	Bardzo niska	Niska	Niezła	Odrobinę poniżej zadowalającej	Zadowalająca	Odrobinę powyżej zadowalającej	Dobra	Bardzo dobra	Świetna	Doskonała

Czytanie

0	1	2	3	4	5	6	7	8	9	10
Zerowa	Bardzo niska	Niska	Niezła	Odrobinę poniżej zadowalającej	Zadowalająca	Odrobinę powyżej zadowalającej	Dobra	Bardzo dobra	Świetna	Doskonała

(4) Według Ciebie, na ile mocny jest u Ciebie angielski akcent, gdy mówisz po polsku?

0	1	2	3	4	5	6	7	8	9	10
Brak	Prawie brak	Bardzo delikatny	Delikatny	W pewnym stopniu	Umiarkowany	Znaczący	Mocny	Bardzo mocny	Wyjątkowo mocny	Dominujący

(5) Proszę zakreśli, jak często inni identyfikują Cię jako nierodzimego użytkownika języka polskiego na podstawie Twojego akcentu, gdy mówisz po polsku:

0	1	2	3	4	5	6	7	8	9	10
Nigdy	Prawie nigdy				W połowie przypadków					Zawsze



Spanish gender-specific analyses

As mentioned in Section 3.7, the effect of gender was found to not be significant. Here I provide the coefficient matrix for the mixed effects model with gender as a fixed effect. For completeness, I provide separate analyses for feminine target gender conditions (Appendix C.1) and for masculine target gender conditions (Appendix C.2).

	<i>Dependent variable:</i>		
	<i>Time of first fixation</i>		
gender	-13.667	(16.908)	
group	72.954	(24.809)	***
condition	-247.737	(33.617)	***
trial number	22.139	(11.674)	*
gender : group	18.700	(22.306)	
gender : condition	49.361	(27.296)	*
group : condition	149.717	(42.008)	***
gender : trial number	-11.735	(16.680)	
group : trial number	-15.431	(15.890)	
condition : trial number	-73.646	(19.239)	***
gender : group : condition	-43.994	(35.372)	
gender : group : trial number	10.541	(22.051)	
gender : condition : trial number	41.345	(27.440)	
group : condition : trial number	28.499	(25.674)	
gender : group : condition : trial number	-35.658	(35.797)	
constant	1,114.137	(19.508)	***
Observations	2,016		
Log Likelihood	-13,420.740		
Akaike Inf. Crit.	26,881.480		
Bayesian Inf. Crit.	26,993.650		

Note: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table C.1: Output of the mixed effects model fitted to the Spanish data for all conditions, with a fixed effect of GENDER. The model also included a random intercept and slope for GENDER and CONDITION conditioned on PARTICIPANT.

C.I FEMININE TARGET ITEM

	<i>Dependent variable:</i>		
	<i>Time of first fixation</i>		
group	72.014	(16.560)	***
condition	85.296	(10.243)	***
trial number	-14.330	(6.350)	**
group : condition	-36.414	(10.243)	***
group : trial number	-1.195	(6.350)	
condition : trial number	28.791	(6.366)	***
group : condition : trial number	-7.202	(6.366)	
constant	1,063.912	(16.560)	***
Observations	996		
Log Likelihood	-6,652.730		
Akaike Inf. Crit.	13,329.460		
Bayesian Inf. Crit.	13,388.300		

Note: *p<0.1; **p<0.05; ***p<0.01

Table C.2: Output of the mixed effects model fitted to the Spanish data for conditions in which the target item was feminine.

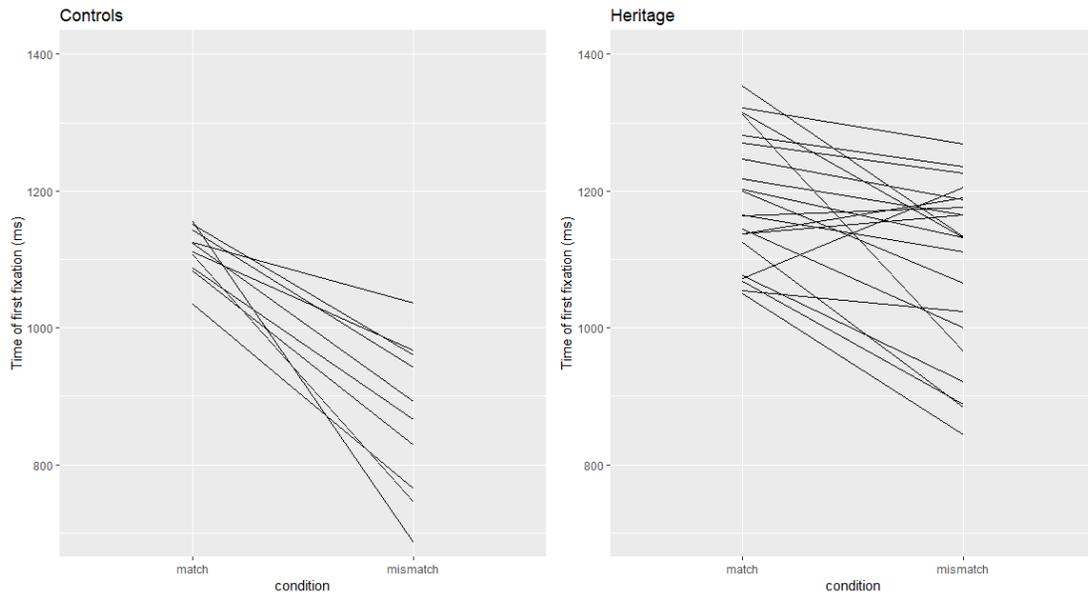


Figure C.1: The time of first fixation on feminine target items across match and mismatch conditions for each participant group.

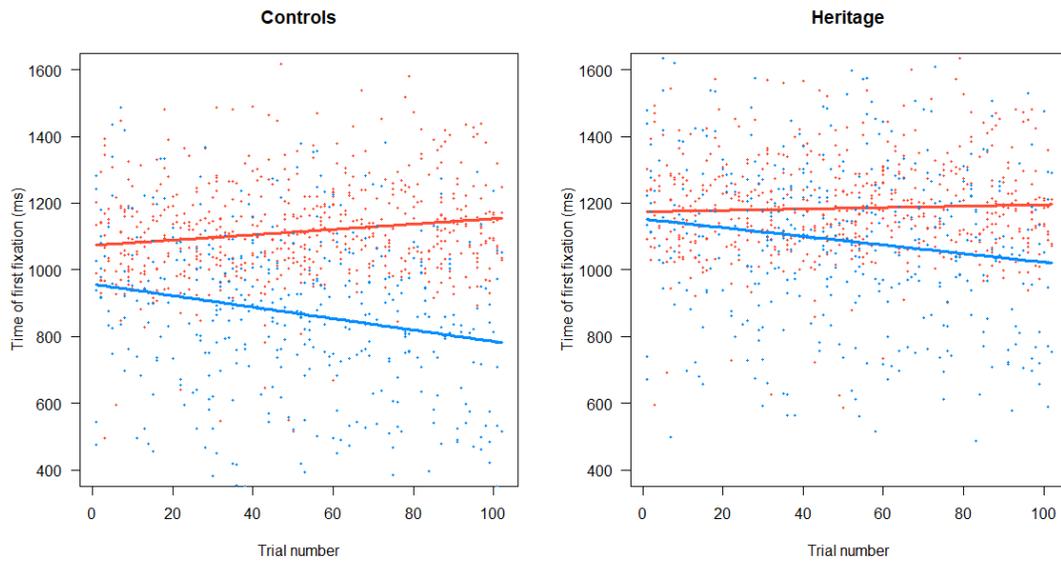


Figure C.2: The time of first fixation on feminine target items across match and mismatch conditions as a function of trial number. The red line indicates the match condition, the blue line indicates the mismatch condition.

C.2 MASCULINE TARGET ITEM

	<i>Dependent variable:</i>		
	<i>Time of first fixation</i>		
group	73.240	(16.570)	***
condition	73.316	(11.248)	***
trial number	-9.643	(6.351)	
group : condition	-26.741	(11.248)	**
group : trial number	-4.378	(6.351)	
condition : trial number	17.151	(6.362)	***
group : condition : trial number	2.462	(6.362)	
constant	1,074.005	(16.570)	***
Observations	1,020		
Log Likelihood	-6,798.224		
Akaike Inf. Crit.	13,620.450		
Bayesian Inf. Crit.	13,679.580		

Note: * p<0.1; ** p<0.05; *** p<0.01

Table C.3: Output of the mixed effects model fitted to the Spanish data for conditions in which the target item was masculine.

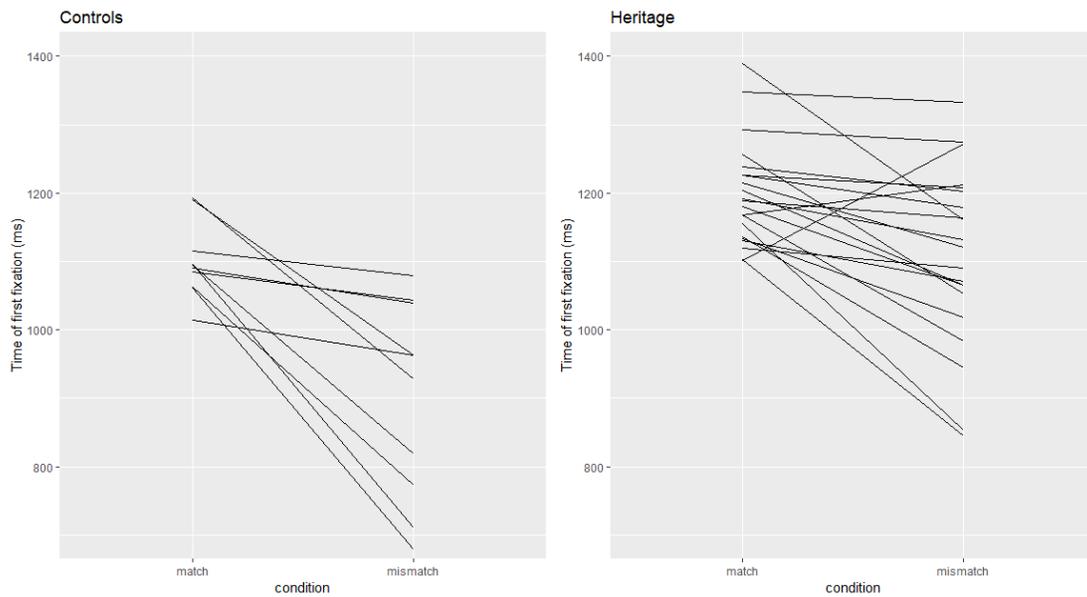


Figure C.3: The time of first fixation on masculine target item across match and mismatch conditions for each participant group.

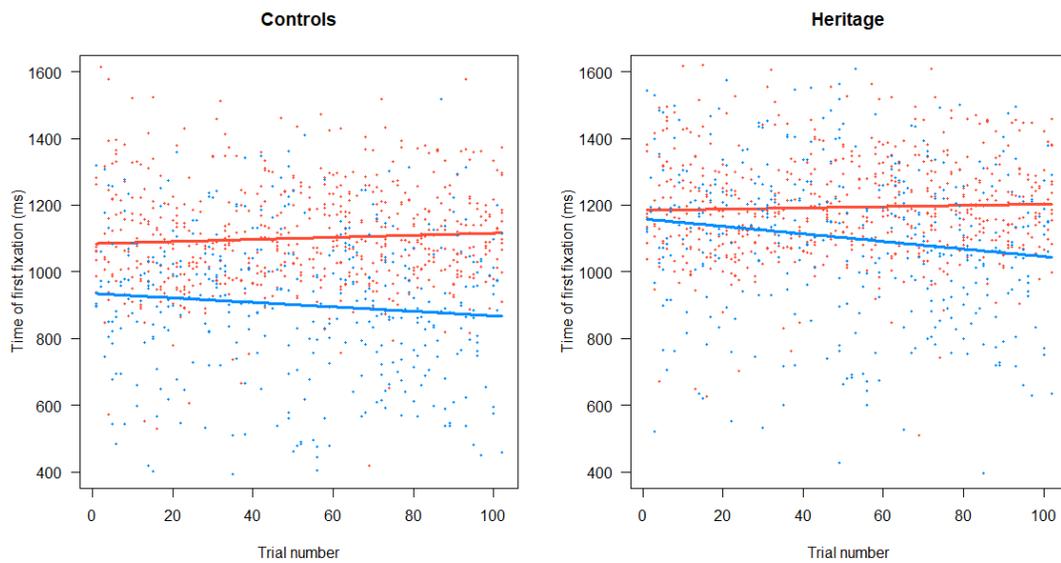


Figure C.4: The time of first fixation on masculine target items across match and mismatch conditions as a function of trial number. The red line indicates the match condition, the blue line indicates the mismatch condition.

References

- Abney, S. P. (1987). *The English Noun Phrase in its Sentential Aspect*. PhD thesis, Massachusetts Institute of Technology.
- Abu-Akel, A., Bailey, A. L., & Thum, Y.-M. (2004). Describing the Acquisition of Determiners in English: A Growth Modeling Approach. *Journal of Psycholinguistic Research*, 33(5), 407–424.
- Ahn, S., Chang, C., DeKeyser, R., & Lee-Ellis, S. (2017). Age Effects in First Language Attrition: Speech Perception by Korean-English Bilinguals. *Language Learning*, 67(3), 694–733.
- Alameda, J. R. & Cuertos, F. (1995). *Diccionario de frecuencias de las unidades lingüísticas del castellano*. Universidad de Oviedo.
- Alarcón, I. (2006). *The second language acquisition of Spanish gender agreement: The effects of linguistic variables on accuracy*. PhD thesis.
- Alarcón, I. (2009). The Processing of Gender Agreement in L1 and L2 Spanish: Evidence from Reaction Time Data. *Hispania*, 92, 814–828.
- Alarcón, I. (2011). Spanish gender agreement under complete and incomplete acquisition: Early and late bilinguals' linguistic behavior within the noun phrase. *Bilingualism: Language and Cognition*, 14(3), 332–350.
- Albirini, A., Benmamoun, E., & Chakrani, B. (2013). Gender and number agreement in the oral production of Arabic Heritage speakers. *Bilingualism: Language and Cognition*, 16(01), 1–18.
- Albirini, A., Benmamoun, E., & Saadah, E. (2011). Grammatical features of Egyptian and Palestinian Arabic heritage speakers' oral production. *Studies in Second Language Acquisition*, 33(02), 273–303.
- Ammerlaan, T. (1996). *You get a bit wobbly...?: exploring bilingual lexical retrieval processes in the context of first language attrition*. PhD thesis, University of Nijmegen.
- Anderson, R. T. (1999). Loss of Gender Agreement in L1 Attrition: Preliminary Results. *Bilingual Research Journal*, 23(4), 389–408.

- Au, T. K.-f., Knightly, L. M., Jun, S.-A., & Oh, J. S. (2002). Overhearing a Language During Childhood. *Psychological Science*, 13(3), 238–243.
- Baker, M. (2008). *The Syntax of Agreement and Concord (Cambridge Studies in Linguistics 115)*. Cambridge: Cambridge University Press.
- Bassano, D., Mailllochon, I., & Mottet, S. (2008). Noun grammaticalization and determiner use in French children's speech: A gradual development with prosodic and lexical influences. *Journal of Child Language*, 35, 403–438.
- Bedore, L. M. & Leonard, L. B. (2001). Grammatical Morphology Deficits in Spanish-Speaking Children With Specific Language Impairment. *Journal of Speech, Language, and Hearing Research*, 44(4), 905–924.
- Behaghel, O. (1909). Beziehungen zwischen Umfang und Reihenfolge von Satzgliedern [Relationships between size and ordering of constituents]. *Indogermanische Forsch*, 25, 110–142.
- Benmamoun, E., Albirini, A., Saadah, E., & Montrul, S. (2008). Agreement and plural features in heritage Arabic speakers. In *Paper presented at the Second Heritage Language Summer Institute, Harvard University, Cambridge, MA*.
- Benmamoun, E., Montrul, S., & Polinsky, M. (2013a). Defining an “ideal” heritage speaker: Theoretical and methodological challenges. Reply to peer commentaries. *Theoretical Linguistics*, 39(3-4), 259–294.
- Benmamoun, E., Montrul, S., & Polinsky, M. (2013b). Heritage languages and their speakers: Opportunities and challenges for linguistics. *Theoretical Linguistics*, 39(3-4), 129–181.
- Bernstein, J. (1993). The syntactic role of word markers in null nominal constructions. *Probus*, 5, 5–38.
- Bialystok, E. (1994). Representation and ways of knowing: Three issues in second language acquisition. In N. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 549–569). San Diego, CA: Academic Press.
- Blake, R. (1983). Mood selection among Spanish-speaking children, ages 4 to 12. *Bilingual Rev.*, 10, 21–32.
- Boberg, C. (1999). The Attitudinal Component of Variation in American English Foreign (a) Nativization. *Journal of Language and Social Psychology*, 18(1), 49–61.
- Boberg, C. (2009). The emergence of a new phoneme: Foreign (a) in Canadian English. *Language Variation and Change*, 21, 355–380.
- Bolonyai, A. (2007). (In)vulnerable agreement in incomplete bilingual L1 learners. *The International Journal of Bilingualism*, 11(1), 3–23.

- Booth, A. & Waxman, S. (2003). Mapping Words to the World in Infancy: Infants' Expectations for Count Nouns and Adjectives. *Journal of Cognition and Development*, 4(3), 357–381.
- Boškovic, Z. (2008). What will you have, DP or NP? In E. Elfner & M. Walkow (Eds.), *Proceedings of the North East Linguistic Society*, volume 37 (pp. 101–114). Amherst: University of Massachusetts.
- Boškovic, Z. (2009). More on the no-DP analysis of article-less languages. *Studia Linguistica*, 63(2), 187–203.
- Bowles, M. A. (2011). Measuring implicit and explicit linguistic knowledge: What Can Heritage Language Learners Contribute? *Studies in Second Language Acquisition*, 33, 247–271.
- Brehmer, B. & Rothweiler, M. (2012). The acquisition of gender agreement marking in Polish. In K. Braunmuller & C. Gabriel (Eds.), *Multilingual Individuals and Multilingual Societies* (pp. 81–100). Amsterdam: John Benjamins.
- Brouwer, S., Sprenger, S., & Unsworth, S. (2017). Processing grammatical gender in Dutch: Evidence from eye movements. *Journal of Experimental Child Psychology*, 159, 50–65.
- Brown, D. & Hippisley, A. (2012). *Network morphology: a defaults-based theory of word structure*. Cambridge: Cambridge University Press.
- Bruening, B. (2009). Selectional Asymmetries between CP and DP Suggest that the DP Hypothesis is Wrong. *Proceedings of the 32nd Annual Penn Linguistics Conference*, 15(1), 27–35.
- Bull, W. (1965). *Spanish for teachers; applied linguistics*. New York: Ronald Press.
- Bylund, E. (2009). Maturation Constraints and First Language Attrition. *Language Learning*, 59(3), 687–715.
- Cantone, K. (2007). *Code-switching in bilingual children*. Dordrecht: Springer.
- Cantone, K. F. (1999). *Das Genus im Italienischen und Deutschen: Empirische Untersuchung zum bilingualen Erstspracherwerb (unpublished master's thesis)*. PhD thesis, University of Hamburg.
- Carreira, M. & Kagan, O. (2011). The Results of the National Heritage Language Survey: Implications for Teaching, Curriculum Design, and Professional Development. *Foreign Language Annals*, 44(1), 40–64.
- Carroll, J. B. (1939). *Determining and Numerating Adjectives in Children's Speech*. Technical Report 4.
- Carroll, S. (1989). Second-Language Acquisition and the Computational Paradigm. *Language Learning*, 39(4), 535–594.

- Carstens, V. (1991). *The morphology and syntax of determiner phrases in Kiswabili*. PhD thesis, UCLA.
- Castilla, A. P. & Pérez-Leroux, A. T. (2010). Language Acquisition Omissions and Substitutions in Spanish Object Clitics: Developmental Optionality as a Property of the Representational System. *Language Acquisition*, 17, 2–25.
- Chang, C. & Yao, Y. (2016). Toward an Understanding of Heritage Prosody: Acoustic and Perceptual Properties of Tone Produced by Heritage, Native, and Second Language Speakers of Mandarin. *The Heritage Language Journal*, 13(12), 134–160.
- Chang, C. & Yao, Y. (2019). Production of neutral tone in Mandarin by heritage, native, and second language speakers. In *Proceedings of the 19th International Congress of Phonetic Sciences* Melbourne, Australia.
- Chomsky, C. (1969). *The acquisition of syntax in children 5 to 10*. Cambridge, MA: MIT Press.
- Chomsky, N. (1965). *Aspects of the theory of syntax*. M.I.T. Press.
- Chomsky, N. (2001). Derivation by Phase. In M. Kenstowicz (Ed.), *Ken Hale: A Life in Language* (pp. 1–52). Cambridge, MA: MIT Press.
- Chung, S. (2013). The Syntactic Relations behind Agreement. In L. Cheng & N. Corver (Eds.), *Diagnosing Syntax*. Oxford: Oxford University Press.
- Clark, E. (1986). Acquisition of Romance, with special reference to French. In D. I. Slobin (Ed.), *The crosslinguistic study of language acquisition: the data (Vol. 1)* (pp. 687–782). Hillsdale, NJ: Erlbaum.
- Comrie, B. (1999). Grammatical Gender Systems: A Linguist's Assessment. *Journal of Psycholinguistic Research*, 28(5), 457–466.
- Cook, V., Ed. (2003). *Effects of the second language on the first*. Multilingual Matters.
- Cook, V. & Newson, M. (2007). *Chomsky's universal grammar : an introduction*. Blackwell Pub.
- Corbett, G. (1983). The Number of Genders in Polish. *Papers and Studies in Contrastive Linguistics*, 16, 83–89.
- Corbett, G. (1991). *Gender*. Cambridge: Cambridge University Press.
- Corbett, G. & Fraser, N. (1999). Default genders. In Barbara Unterbeck & Matti Rissanen (Eds.), *Gender in Grammar and Cognition* chapter Default ge, (pp. 55–97). Berlin: Mouton de Gruyter.
- Corver, N. (2006). Subextraction. In M. Everaert & H. van Riemsdijk (Eds.), *The Blackwell Companion to Syntax (Volume IV)* chapter 66, (pp. 566–600). Malden, MA: Blackwell.

- Curcic, M., Andringa, S., & Kuiken, F. (2018). The Role of Awareness and Cognitive Aptitudes in L2 Predictive Language Processing. *Language Learning*, 69, 42–71.
- Dabrowska, E. (2006). Low-level schemas or general rules? The role of diminutives in the acquisition of Polish case inflections. *Language Sciences*, 28, 120–135.
- De Groot, C. (2005). The grammars of Hungarian outside Hungary from a linguistic-typological perspective. In A. Fenyvesi (Ed.), *Hungarian Language Contact Outside Hungary* (pp. 351–370). Amsterdam: John Benjamins.
- de La Grasserie, R. (1898). *Langue zoque et langue mixte; grammaire, dictionnaire*. Paris: J. Maisonneuve.
- Dominguez, A., Cuetos, F., & Segui, J. (1999). The Processing of Grammatical Gender and Number in Spanish. *Journal of Psycholinguistic Research*, 28(5).
- Dubinina, I. (2012). *How to Ask for a Favor: An Exploration of Speech Act Pragmatics in Heritage Russian*. PhD thesis, Bryn Mawr College.
- Dukiewicz, L., Urbańczyk, S., & Polskiego, I. J. (1984). *Gramatyka współczesnego języka polskiego*. Warsaw: PWN.
- Dussias, P., Beatty-Martinez, A., & Perrotti, L. (2017). Susceptibility to interference affects the second and the first language. *Bilingualism: Language and Cognition*, 20(4), 681–682.
- Dussias, P. E., Valdés Kroff, J. R., Guzzardo Tamargo, R. E., & Gerfen, C. (2013). When gender and looking go hand in hand. *Studies in Second Language Acquisition*, 35(2), 353–387.
- Eichler, N. (2011). *Code-Switching bei bilingual aufwachsenden Kindern: Eine Analyse der gemischtsprachlichen Nominalphrasen unter besonderer Berücksichtigung des Genus [Code-Switching in bilingual children: An analysis of mixed DPs and the role of gender]*. PhD thesis, Narr. Bergische Universität Wuppertal, Germany.
- Eichler, N., Hager, M., & Müller, N. (2012). CODE-SWITCHING WITHIN DETERMINER PHRASES IN BILINGUAL CHILDREN : French , Italian , Spanish and German Author (s): Nadine Eichler , Malin Hager and Natascha Müller Source : Zeitschrift für französische Sprache und Literatur , Bd . 122 , H . 3 (2012),. *Zeitschrift für französische Sprache und Literatur*, 122(3), 227–258.
- Eichler, N., Jansen, V., & Müller, N. (2013). Gender acquisition in bilingual children: French–German, Italian–German, Spanish–German and Italian–French. *International Journal of Bilingualism*, 17(5), 550–572.
- Ellis, R. (2005). Measuring implicit and explicit knowledge of a second language: A Psychometric Study. *Studies in Second Language Acquisition*, 27, 141–172.

- Ellis, R., Loewen, S., & Erlam, R. (2006). Implicit and explicit corrective feedback and the acquisition of L2 grammar. *Studies in Second Language Acquisition*, 28, 339–368.
- Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., Tomasello, M., Mervis, C. B., & Stiles, J. (1994). *Variability in Early Communicative Development*. Technical Report 5.
- Fenyvesi, A. (2000). The affectedness of the verbal complex in American Hungarian. In A. Fenyvesi & K. Sándor (Eds.), *Language contact and the verbal complex of Dutch and Hungarian: Working papers from the 1st Bilingual Language Use Theme Meeting of the Study Centre on Language Contact, November 11-13, 1999, Szeged, Hungary* (pp. 94–107). Szeged: JGyTF Press.
- Fishman, J. (2006). Three-hundred plus years of heritage language education in the United States. In G. Valdés, J. Fishman, R. Chavez, & W. Perez (Eds.), *Developing minority language resources. The case of Spanish in California* (pp. 12–23). Clevedon: Multilingual Matters.
- Franceschina, F. (2001). Morphological or syntactic deficits in near-native speakers? An assessment of some current proposals. *Second Language Research*, 17(3), 213–247.
- Friederici, A. D. & Jacobsen, T. (1999). *Processing Grammatical Gender During Language Comprehension*. Technical Report 5.
- Fuchs, Z. (2014). Gender and Analogical Extension: From animacy to borrowings in Polish. *New Insights into Slavic Linguistics*, 3, 115–127.
- Fuchs, Z., Polinsky, M., & Scontras, G. (2015). The differential representation of number and gender in Spanish. *Linguistic Review*, 32(4), 703–737.
- Gabor, B. & Lukacs, A. (2012). Early morphological productivity in Hungarian: evidence from sentence repetition and elicited production. *Journal of Child Language*, (pp. 411–442).
- Gass, S. & Selinker, L. (1992). *Language Transfer in Language Learning*, volume 5 of *Language Acquisition and Language Disorders*. Amsterdam: John Benjamins Publishing Company.
- Godson, L. (2003). *Phonetics of language attrition: Vowel production and articulatory setting in the speech of Western Armenian heritage speakers*. PhD thesis, University of California, San Diego.
- Godson, L. (2004). Vowel production in the speech of Western Armenian heritage speakers. *The Heritage Language Journal*, 2(1).
- Gollan, T. H., Slattery, T. J., Goldenberg, D., Van Assche, E., Duyck, W., & Rayner, K. (2011). Frequency Drives Lexical Access in Reading but Not in Speaking: The Frequency-Lag Hypothesis. *Journal of Experimental Psychology: General*, 140(2), 186–209.
- Grosjean, F. (2008). *Studying bilinguals*. Oxford University Press.

- Grüter, T., Hurtado, N., & Fernald, A. (2012). Interpreting Object Clitics in Real Time: Eye-Tracking Evidence from 4-Year-Old and Adult Speakers of Spanish. *Proceedings of the Annual Boston University Conference on Language Development*, 36(Dc 008838), 213–225.
- Grüter, T., Lew-Williams, C., & Fernald, A. (2012). Grammatical gender in L2: A production or a real-time processing problem? *Second Language Research*, 28(2), 191–215.
- Guasti, M. T. (2002). *Language acquisition : the growth of grammar*. MIT Press.
- Haiman, J. (1983). Iconic and Economic Motivation. *Language*, 59(4), 781–819.
- Håkansson, G. (1995). Syntax and morphology in language attrition: a study of five bilingual expatriate Swedes. *International Journal of Applied Linguistics*, 5(2), 153–169.
- Halle, M. & Matushansky, O. (2006). The Morphophonology of Russian Adjectival Inflection. *Linguistic Inquiry*, 37(3), 351–404.
- Haman, E., Etenkowski, M., Łuniewska, M., Szwabe, J., Dąbrowska, E., Szreder, M., & Łaziński, M. (2011). *Polish CDS Corpus*. Technical report, Talkbank.
- Harley, H. (2013). Squib: Feature-matching and case and number dissociation in Hiaki. *Sintaxe e semântica formais*, 9(1), 1–10.
- Harris, J. W. (1985). Spanish Word Markers. In F. H. Nuessel Jr. (Ed.), *Current Issues in Spanish Phonology and Morphology*. Bloomington, IN: Indiana University Linguistics Club Publications.
- Harris, J. W. (1991). The Exponence of Gender in Spanish. *Linguistic Inquiry*, 22(1), 27–62.
- Haspelmath, M. (2006). Against Markedness (And What to Replace It With) Author (s): Martin Haspelmath Published by : Cambridge University Press Stable URL : <http://www.jstor.org/stable/4176968> Accessed : 30-05-2016 19 : 58 UTC Your use of the JSTOR archive indicates your ac. 42(1), 25–70.
- Heine, B. (1968). The allocation of loan-words within the nominal class systems of some Togo Remnant Languages. *African Languages*, 7, 130.
- Henderson, B. (2006). Multiple Agreement, Concord and Case Checking in Bantu. In O. Arasanyin & M. Pemberton (Eds.), *Selected Proceedings of the 36th Annual Conference on African Linguistics* (pp. 60–65). Somerville, MA: Cascadilla Proceedings Project.
- Hernandez Pina, F. (1984). *Teorías psicosociolingüísticas y su aplicación a la adquisición del español como lengua materna*. Madrid: Siglo XXI.
- Hockett, C. (1958). *A course in modern linguistics*. New York: Macmillan.
- Hopp, H. (2013). Grammatical gender in adult L2 acquisition: Relations between lexical and syntactic variability. *Second Language Research*, 29(1), 33–56.

- Hopp, H. (2016). Learning (not) to predict: Grammatical gender processing in second language acquisition. *Second Language Research*, 32(2), 277–307.
- Hopp, H. & Lemmerth, N. (2018). Lexical and syntactic congruency in L2 predictive gender processing. *Studies in Second Language Acquisition*, 40(1), 171–199.
- Howard, C. & Holcombe, A. (2010). Unexpected changes in direction of motion attract attention. *Attention, Perception, & Psychophysics*, 72(8), 2087–2095.
- Huettig, F. & Altmann, G. T. M. (2011). Looking at anything that is green when hearing "frog": How object surface colour and stored object colour knowledge influence language-mediated overt attention. *The Quarterly Journal of Experimental Psychology*, 64(1), 122–145.
- Huettig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica*, 137(2), 151–171.
- Hulsen, M. (2000). *Language loss and language processing : three generations of Dutch migrants in New Zealand*. PhD thesis, University of Nijmegen.
- Ionin, T. & Montrul, S. (2010). The Role of L1 Transfer in the Interpretation of Articles with Definite Plurals in L2 English. *Language Learning*, 60(4), 877–925.
- Isurin, L. & Ivanova-Sullivan, T. (2008). Lost in between: the case of Russian heritage speakers. *The Heritage Language Journal*, 6, 72–104.
- Ivanova-Sullivan, T. (2014). *Theoretical and Experimental Aspects of Syntax-Discourse Interface in Heritage Grammars*. Brill.
- Janssen, B. (2014). Frequency Effects on the Acquisition of Polish and Russian Gender Morphology. *Studies in Slavic and General Linguistics*, 40, 109–126.
- Janssen, B. (2016). *The acquisition of gender and case in Polish and Russian : a study of monolingual and bilingual children*. Amsterdam: Uitgeverij Pegasus.
- Jarvis, S. (1998). *Conceptual transfer in the interlingual lexicon*. Bloomington, IN: Indiana University Linguistics Club Publications.
- Kang, H.-S. & Kim, I.-S. (2012). Perceived and actual competence and ethnic identity in heritage language learning: a case of Korean-American college students. *International Journal of Bilingual Education and Bilingualism*, 15(3), 279–294.
- Karmiloff-Smith, A. (1979). *A functional approach to child language : A study of determiners and reference (Cambridge Studies in Linguistics 24)*. Cambridge, MA: Cambridge University Press.
- Keating, G. D. (2009). Sensitivity to violation of gender agreement in native and nonnative Spanish. *Language Learning*, 59(3), 503 – 535.

- Kempe, V. & Brooks, P. J. (2001). The role of diminutives in Russian gender learning: Can child-directed speech facilitate the acquisition of inflectional morphology. *Language Learning*, 51(2), 221–256.
- Kim, J.-h. (2007). *Binding interpretations in Korean heritage speakers and L2 learners*. Dissertation, University of Illinois at Urbana-Champaign.
- Klemensiewicz, Z. (1960). *Podstawowe wiadomości z gramatyki języka polskiego*. Warsaw: PWN, 8 edition.
- Klockmann, H. (2012). *Polish numerals and quantifiers : A syntactic analysis of subject - verb agreement mismatches*. Master's thesis, Utrecht University.
- Knightly, L. M., Jun, S.-A., Oh, J. S., & Kit-Fong Au, T. (2003). Production benefits of childhood overhearing. *The Journal of the Acoustical Society of America*, 114, 465.
- Koolen, R., Goudbeek, M., & Krahmer, E. (2013). The Effect of Scene Variation on the Redundant Use of Color in Definite Reference. *Cognitive Science*, 37(2), 395–411.
- Krajewski, G. (2005). *The role of grammatical gender in the acquisition of noun inflection in Polish*. Technical Report 1.
- Kramer, R. (2014). Gender in Amharic: a morphosyntactic approach to natural and grammatical gender. *Language Sciences*, 43, 102–115.
- Kramer, R. (2015). *The Morphosyntax of Gender*. Oxford University Press.
- Kuchenbrandt, I. (2005). Gender acquisition in bilingual Spanish. *Proceedings of the 4th International Symposium on ...*, (April), 1252–1263.
- Kupisch, T., Akpınar, D., & Stöhr, A. (2013). Gender assignment and gender agreement in adult bilinguals and second language learners of French. *Linguistic Approaches to Bilingualism*, 3(2), 150–179.
- Kupisch, T. & Rothman, J. (2018). Terminology matters! Why difference is not incompleteness and how early child bilinguals are heritage speakers. *International Journal of Bilingualism*, 22(5), 564–582.
- Lee, S.-H., Bong, M., & Polinsky, M. (2012). *Nouns and verbs in heritage Korean*. Manuscript, Wellesley and Harvard University.
- Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22, 1–75.
- Lew-Williams, C. & Fernald, A. (2007). Young Children Learning Spanish Make Rapid Use of Grammatical Gender in Spoken Word Recognition. *Psychological Science*, 18(3), 193–198.

- Lew-Williams, C. & Fernald, A. (2010). Real-time processing of gender-marked articles by native and non-native Spanish speakers. *Journal of Memory and Language*, 63(4), 447–464.
- Liceras, J., Fernández Fuertes, R., Perales, S., Pérez-Tattam, R., & Spradlin, K. T. (2008). Gender and gender agreement in bilingual native and non-native grammars: A view from child and adult functional-lexical mixings. *Lingua*, 118, 827–851.
- Lleó, C. (1997). Filler syllables, proto-articles and early prosodic constraints in Spanish and German. In *Proceedings of GALA '97 Conference on Language Acquisition* (pp. 251–256). Edinburgh: The University of Edinburgh.
- Lleó, C. (1998). Proto-articles in the acquisition of Spanish: Interface between Phonology and Morphology. In R. Fabri, A. Ortmann, & T. Parodi (Eds.), *Modelle der Flexion: 18. Jahrestagung der Deutschen Gesellschaft für Sprachwissenschaft* (pp. 175–195). Tübingen: Niemeyer.
- Loerts, H. (2012). *Eyes and brains, native and second language learners, & grammatical gender*. PhD thesis, University of Groningen.
- Loerts, H., Wieling, M., & Schmid, M. S. (2013). Neuter is not Common in Dutch: Eye Movements Reveal Asymmetrical Gender Processing. *Journal of Psycholinguistic Research*, 42, 551–570.
- Lohndal, T. (2013). Generative grammar and language mixing. *Theoretical Linguistics*, 39(3-4), 215–224.
- Longobardi, G. (1994). Reference and Proper Names: A Theory of N-Movement in Syntax and Logical Form. *Linguistic Inquiry*, 25(4), 609–665.
- Longobardi, G. (1996). *The syntax of n-raising: a minimalist theory*. Utrecht: Utrecht Univ. Research Institute for Language and Speech.
- López-Ornat, S. (1997). What Lies in between a Pre-Grammatical and a Grammatical Representation? Evidence on Nominal and Verbal Form-Function Mappings in Spanish from 1;7 to 2;1. In A. T. Perez-Leroux & W. R. Glass (Eds.), *Contemporary Perspectives on the Acquisition of Spanish, Vol 1: Developing Grammars* (pp. 3–20). Somerville, MA: Cascadilla.
- López-Ornat, S., Fernández, A., Gallo, P., & Mariscal, S., Eds. (1994). *La adquisición de la lengua española*. Madrid: Siglo XXI.
- Luczynski, E. (2002). Zagadnienie kompetencji fleksyjnej dzieci w wieku przedszkolnym [na przykładzie rzeczownika] (The issue of inflectional competence of preschool children [case of noun]). *Jezyk Polski*, 1, 43–50.
- MacWhinney, B. (1978). The Acquisition of Morphophonology. *Monographs of the Society for Research on Child Development*, 43, 1–123.

- Mahajan, G. (2009). Ongoing deficits in heritage Hindi. In *Paper presented at the Third Annual Heritage Language Institute, Urbana Champaign, June 2009*.
- Mańczak, W. (1956). Ile rodzajów jest w polskim? *Język polski*, 36(2), 112–121.
- Mani, N., Johnson, E., McQueen, J. M., & Huettig, F. (2013). How yellow is your banana? toddlers' language-mediated visual search in referent-present tasks. *Developmental Psychology*, 49(6), 1036–1044.
- Maratsos, M. (1988). The acquisition of formal word classes. In Y. Levy, I. Schlesinger, & M. Braine (Eds.), *Categories and processes in language acquisition*. Hillsdale, NJ: Erlbaum.
- Mariscal, S. (1996). Adquisiciones morfosintacticas en torno al sintagma nominal: El genero gramatical en espanol. In M. Pérez-Pereira (Ed.), *Estudios sobre la adquisicion del Castellano, Catalan, Euskera y Gallego* (pp. 263–272). Universidade de Santiago de Compostela, Servicio de Publicaciones.
- Mariscal, S. (2001). Es 'a pie' equivalente a DET + N? Sobre el conocimiento temprano de las categorias gramaticales. *Cognitiva*, 13(1), 35–59.
- Mariscal, S. (2009). Early acquisition of gender agreement in the Spanish noun phrase: starting small. *Journal of Child Language*, 36, 143–171.
- Marrero, V., Aguirre, C., & Albala, M. (2007). The acquisition of diminutives in Spanish. In I. Savickienė & W. Dressler (Eds.), *The acquisition of diminutives: A crosslinguistic perspective* (pp. 155–184). Amsterdam: John Benjamins.
- Martinez-Mira, M.-I. (2009). Spanish heritage speakers in the Southwest: Factors contributing to the maintenance of the subjunctive in concessive clauses. *Spanish in Context*, 6(1), 105–126.
- Menyuk, P. & Brisk, M. E. (2005). *Language Development and Education*. Palgrave Macmillan.
- Merino, B. (1983). Language loss in bilingual Chicano children. *Journal of Applied Developmental Psychology*, 4(3), 277–294.
- Mintz, T. H. (2003). Frequent frames as a cue for grammatical categories in child directed speech. *Cognition*, 90, 91–117.
- Moag, R. (1995). Semi-native speakers: how to hold and mold them. In V. Gambhir (Ed.), *The Teaching and Acquisition of South Asian Languages* (pp. 168–181). Philadelphia: University of Pennsylvania Press.
- Montrul, S. (2002). Incomplete acquisition and attrition of Spanish tense/ aspect distinctions in adult bilinguals. *Bilingualism: Language and Cognition*, 5(1), 39–68.

- Montrul, S. (2004). Subject and object expression in Spanish heritage speakers: A case of morphosyntactic convergence. *Bilingualism: Language and Cognition*, 7(2), 125–142.
- Montrul, S. (2005). Second language acquisition and first language loss in adult early bilinguals: exploring some differences and similarities. *Second Language Research*, 21, 199–249.
- Montrul, S. (2007). Interpreting mood distinctions in Spanish as a heritage language. In K. Potowski & R. Cameron (Eds.), *Spanish in Contact. Policy, Social and Linguistic Inquiries* (pp. 23–40). Amsterdam: John Benjamins.
- Montrul, S. (2008). *Incomplete Acquisition in Bilingualism. Re-examining the Age Factor*. Amsterdam: John Benjamins.
- Montrul, S. (2009). Knowledge of tense-aspect and mood in Spanish heritage speakers. *International Journal of Bilingualism*, 13(2), 239–269.
- Montrul, S. (2010a). Current issues in heritage language acquisition. *Annual Review of Applied Linguistics*, 30, 3–23.
- Montrul, S. (2010b). How similar are adult second language learners and Spanish heritage speakers? Spanish clitics and word order. *Applied Psycholinguistics*, 31, 167–207.
- Montrul, S. (2016). *The Acquisition of Heritage Languages*. Cambridge: Cambridge University Press.
- Montrul, S., Bhatt, R. M., & Bhatia, A. (2012). Erosion of case and agreement in Hindi heritage speakers. *Linguistic Approaches to Bilingualism*, 2(2), 141–176.
- Montrul, S. & Bowles, M. (2009). Back to basics: Incomplete knowledge of Differential Object Marking in Spanish heritage speakers. *Bilingualism: Language and Cognition*, 12(03), 363.
- Montrul, S., Davidson, J., de la Fuente, I., & Foote, R. (2014). Early language experience facilitates the processing of gender agreement in Spanish heritage speakers. *Bilingualism: Language and Cognition*, 17(01), 118–138.
- Montrul, S., de la Fuente, I., Davidson, J., & Foote, R. (2013). The role of experience in the acquisition and production of diminutives and gender in Spanish: Evidence from L2 learners and heritage speakers. *Second Language Research*, 29(1), 87–118.
- Montrul, S., Foote, R., & Perpiñán, S. (2008). Gender agreement in adult second language learners and Spanish heritage speakers: The effects of age and context of acquisition. *Language Learning*, 58(3), 503–553.
- Montrul, S. & Ionin, T. (2012). Dominant Language Transfer in Spanish Heritage Speakers and Second Language Learners in the Interpretation of Definite Articles. *The Modern Language Journal*, 96(1), 70–94.

- Montrul, S. & Potowski, K. (2007). Command of gender agreement in school-age Spanish-English Bilingual Children. *The International Journal of Bilingualism / International Journal of Bilingualism*, 11(3), 301–328.
- Montrul, S. & Sánchez-Walker, N. (2013). Differential Object Marking in Child and Adult Spanish Heritage Speakers. *Language Acquisition*, 20(2), 109–132.
- Morales, L., Paolieri, D., Dussias, P. E., Valdés Kroff, J. R., Gerfen, C., & Bajo, M. T. (2016). The gender congruency effect during bilingual spoken-word recognition. *Bilingualism*, 19(2), 294–310.
- Mueller Gathercole, V. C. (2002). Grammatical Gender in Bilingual and Monolingual Children: A Spanish Morphosyntactic Distinction. In R. Oller & R. Eilers (Eds.), *Language and Literacy in Bilingual Children* (pp. 207–219). Clevedon: Multilingual Matters.
- Müller, N. (1987). *Der Genuserwerb im Französischen und Deutschen: Eine empirische Untersuchung eines bilingualen Kindes (unpublished master's thesis)*. PhD thesis, University of Hamburg.
- Myers-Scotton, C. & Jake, J. L. (2011). Explaining aspects of code-switching and their implications. In J. L. Nicol (Ed.), *One Mind, Two Languages: Bilingual Language Processing* (pp. 84–116). Maiden: Blackwell.
- Nicenboim, B. & Vasishth, S. (2016). Statistical methods for linguistic research: Foundational Ideas - Part II. *Language and Linguistics Compass*, 10, 591–613.
- Nippold, M. (1998). *Later language development: The school-age and adolescent years (2nd Ed.)*. Austin, TX: Pro-Ed.
- Norris, M. (2014). *A Theory of Nominal Concord*. PhD thesis, University of California, Santa Cruz.
- Odlin, T. (1989). *Language Transfer*. Cambridge: Cambridge University Press.
- O'Grady, W. (1997). *Syntactic Development*. Chicago: The University of Chicago Press.
- O'Grady, W., Lee, M., & Choo, M. (2001). The acquisition of relative clauses by heritage and non-heritage learners of Korean as a second language: A comparative study. *Journal of Korean Language Education*, 12, 283–294.
- O'Grady, W., Lee, M., & Kwak, H.-Y. (2009). Emergentism and second language acquisition. In W. Ritchie & T. Bhatia (Eds.), *Handbook of Second Language Acquisition* (pp. 68–81). Bingley: Emerald Press.
- Oh, J. S., Jun, S.-A., Knightly, L. M., & Kit-Fong Au, T. (2003). Holding on to childhood language memory. *Cognition*, 86, B53–B64.

- Pallier, C. (2007). Critical periods in language acquisition and language attrition. In B. Kopke, M. S. Schmid, M. Keijzer, & S. Dosterst (Eds.), *Language Attrition. Theoretical Perspectives* (pp. 99–120). Amsterdam: John Benjamins.
- Paris, G., Weber, A., & Crocker, M. (2006). No Title. In *Poster presented at the 12th annual conference on architectures and mechanisms for language processing (AMLaP 2006)* Nijmegen, Netherlands.
- Pascual y Cabo, D. (2013). *Agreement reflexes of emerging optionality in heritage speaker Spanish*. PhD thesis, University of Florida.
- Pavlenko, A. & Jarvis, S. (2002). Bidirectional Transfer. *Applied Linguistics*, 23(2), 190–214.
- Pérez-Pereira, M. (1991). The acquisition of gender: What Spanish children tell us. *Journal of Child Language*, 18(3), 571–590.
- Peristeri, E., Tsimpli, I. M., Sorace, A., & Tsapkini, K. (2018). Language interference and inhibition in early and late successive bilingualism. *Bilingualism: Language and Cognition*, 21(5), 1009–1034.
- Piantadosi, S. T., Tily, H., & Gibson, E. . (2012). The Communicative Function of Ambiguity in Language. *Cognition*, 122(3), 280–91.
- Piantadosi, S. T., Tily, H., & Gibson, E. (2018). *Information content versus word length in natural language: A reply to Ferrer-i-Cancho and Moscoso del Prado Martin (2011)*. Technical report.
- Picallo, M. C. (1991). Nominals and Nominalizations in Catalan. *Probus*, 3(3), 279–316.
- Pine, J. M. & Lieven, E. V. M. (1997). *Slot and frame patterns and the development of the determiner category*. Technical report.
- Pinker, S. (1984). *Language Learnability and Language Development*. Cambridge, MA: Harvard University Press.
- Polinsky, M. (1997). Cross-linguistic parallels in language loss. *Southwest Journal of Linguistics*, 14(1-2), 87–123.
- Polinsky, M. (2000). A composite linguistic profile of a speaker of Russian in the U.S. In O. Kagan & B. Rifkin (Eds.), *The Learning and Teaching of Slavic Languages and Cultures: Toward the 21st Century* (pp. 437–65). Bloomington, IN: Slavica.
- Polinsky, M. (2005). Word Class Distinctions in an Incomplete Grammar. In *Perspectives on Language and Language Development* (pp. 419–434). Springer US.
- Polinsky, M. (2006). Incomplete Acquisition: American Russian. *Journal of Slavic Linguistics*, 14(2), 191–262.

- Polinsky, M. (2008a). Heritage language narratives. In D. Brinton, O. Kagan, & S. Bauckus (Eds.), *Heritage language education: A new field emerging* (pp. 149–164). New York, NY: Routledge.
- Polinsky, M. (2008b). Relative Clauses in Heritage Russian: Fossilization or Divergent Grammar? *Formal approaches to Slavic linguistics #16: The Stony Brook meeting 2007*, (pp. 5–21).
- Polinsky, M. (2008c). Russian gender under incomplete acquisition. *The Heritage Language Journal*, 6(1).
- Polinsky, M. (2011). Reanalysis in adult heritage language: New Evidence in Support of Attrition. *Studies in Second Language Acquisition*, 33, 305–328.
- Polinsky, M. (2018). *Heritage Languages and Their Speakers*. Cambridge University Press.
- Poplack, S. (1980). Sometimes I'll start a sentence in Spanish y termino en Español: Toward a typology of code-switching. *Linguistics*, 18(7-8), 581–618.
- Potowski, K., Jegerski, J., & Morgan-Short, K. (2009). The Effects of Instruction on Linguistic Development in Spanish Heritage Language Speakers. *Language Learning*, 59(3), 537–579.
- Prado, M. (1982). El genero en español y la teoria de la marcadez. *Source: Hispania*, 65(2), 258–266.
- Rakhilina, R. & Marushkina, A. (2014). Non-calquing strategies in Heritage Russian.
- Rebuschat, P. & Williams, J. N. (2012). Implicit and explicit knowledge in second language acquisition. *Applied Psycholinguistics*, 33, 829–856.
- Ritter, E. (1991). Two Functional Categories in Noun Phrases: Evidence from Modern Hebrew. *Syntax and Semantics*, 25, 37–62.
- Ritter, E. (1993). Where's Gender? *Linguistic Inquiry*, 24, 795–803.
- Roca, I. M. (1989). The Organization of Grammatical Gender. *Transactions of the Philological Society*, 87, 1–32.
- Rodina, Y. & Westergaard, M. (2012). A cue-based approach to the acquisition of grammatical gender in Russian. *Journal of Child Language*, 39, 1077–1106.
- Rothman, J. (2007). Heritage speaker competence differences, language change, and input type: Inflected infinitives in Heritage Brazilian Portuguese. *The International Journal of Bilingualism 'International Journal of Bilingualism'*, 11(4), 359–389.
- Rothman, J. (2009). Understanding the nature and outcomes of early bilingualism: Romance languages as heritage languages. *International Journal of Bilingualism*, 13(2), 155–163.
- Sagarra, N. & Herschensohn, J. (2011). Proficiency and Animacy Effects on L2 Gender Agreement Processes During Comprehension. *Language Learning*, 61(1), 80–116.

- Saloni, Z. & Świdziński, M. (2001). *Składnia współczesnego języka polskiego*. Warsaw: PWN, 2nd rev. edition.
- Sanchez-Sadek, C. & Others (1975). *The Acquisition of the Concept of Grammatical Gender in Monolingual and Bilingual Speakers of Spanish*.
- Sánchez-Walker, N. (2012). *Comprehension of subject and object relative clauses in Spanish heritage speakers and L2 learners of Spanish*. Phd qualifying paper, University of Illinois at Urbana-Champaign.
- Savickienė, I. & Dressler, W., Eds. (2007). *The Acquisition of Diminutives: A Crosslinguistic Perspective*. Amsterdam: John Benjamins.
- Schenker, A. M. (1955). Linguistic Society of America Gender Categories in Polish. *Language*, 31(3), 402–408.
- Schwartz, B. D. (2004). Why child L2 acquisition? In J. Van Kampen & S. Baaui (Eds.), *Proceedings of Generative Approaches to Language Acquisition* (pp. 47–66). Utrecht, the Netherlands: LOT Occasional Series.
- Schwartz, B. D. & Sprouse, R. A. (1996). *L2 cognitive states and the Full Transfer/Full Access model*. Technical Report 1.
- Scontras, G., Polinsky, M., & Fuchs, Z. (2018). In support of representational economy: Agreement in heritage Spanish. *Glossa: a journal of general linguistics*, 3(1), 1.
- Scontras, G., Polinsky, M., Tsai, C.-Y. E., & Mai, K. (2017). Cross-linguistic scope ambiguity: When two systems meet. *Glossa: a journal of general linguistics*, 2(1), 36.
- Sekerina, I. A. (2015). Predictions, fast and slow. *Linguistic Approaches to Bilingualism*, 5(4), 532–536.
- Seliger, H. (1996). Primary language attrition in the context of bilingualism. In W. Ritchie & T. Bhatia (Eds.), *Handbook of Second Language Acquisition* (pp. 605–625). New York, NY: Academic Press.
- Seva, N., Kempe, V., Brooks, P. J., Mironova, N., & Pershukova, A. (2007). Crosslinguistic evidence for the diminutive advantage: gender agreement in Russian and Serbian children. *Journal of Child Language*, 34, 111–131.
- Silva-Corvalán, C. (1994). *Language contact and change : Spanish in Los Angeles*. Oxford University Press.
- Silva-Corvalán, C. (2003). Linguistic Consequences of Reduced Input in Bilingual First Language Acquisition. In S. Montrul & F. Ordóñez (Eds.), *Linguistic Theory and Language Development in*

- Hispanic Languages: Papers from the 5th Hispanic Linguistics Symposium and the 4th Conference on the Acquisition of Spanish and Portuguese* (pp. 375–397). Somerville, MA: Cascadilla.
- Smoczynska, M. (1985). The acquisition of Polish. In D. Slobin (Ed.), *The crosslinguistic study of language acquisition (Vol. 1)* (pp. 595–686). Hillsdale, NJ: Erlbaum.
- Snyder, W., Senghas, A., & Inman, K. (2001). Agreement Morphology and the Acquisition of Noun-Drop in Spanish. *Language Acquisition*, 9(2), 157–173.
- Soler, R. (1984). Adquisición y utilización del artículo. In M. Siguan (Ed.), *Estudios sobre psicología del lenguaje infantil*. Madrid: Piramide.
- Song, M., O’Grady, W., Cho, S., & Lee, M. (1997). The Learning and Teaching of Korean in Community Schools. *The Korean Language in America*, 2, 111–127.
- Sorace, A. (2004). Native language attrition and developmental instability at the syntax-discourse interface: Data, interpretations and methods. *Bilingualism: Language and Cognition*, 7(2), 143–145.
- Sorace, A. (2011). Pinning down the concept of “interface” in bilingualism. *Linguistic Approaches to Bilingualism*, 1(1), 1–33.
- Swan, O. (2015). Polish Gender, Subgender, and Quasi-Gender. *Journal of Slavic Linguistics*, 23(1), 83–122.
- Szabolcsi, A. (1983). The possessor that ran away from home. *The Linguistic Review*, 3, 89–102.
- Szober, S. (1963). *Gramatyka języka polskiego*. Warsaw: PWN.
- Tomasello, M. (2000). The item-based nature of children’s early syntactic development. *Trends in Cognitive Sciences*, 4(4), 156–163.
- Unsworth, S. (2005). *Child L2, adult L2, child L1 differences and similarities. A study on the acquisition of object scrambling in Dutch*. PhD thesis, Netherlands Graduate School of Linguistics.
- Unsworth, S., Argyri, F., Cornips, L., Hulk, A., Sorace, A., & Tsimpli, I. (2011). Bilingual Acquisition of Greek Voice Morphology and Dutch Gender: What Do They Have in Common? In *Proceedings of the 35th annual Boston University Conference on Language Development* (pp. 590–602).: Cascadilla Press.
- Unsworth, S., Argyri, F., Cornips, L., Sorace, A., & Tsimpli, I. (2014). The role of age of onset and input in early child bilingualism in Greek and Dutch. *Applied Psycholinguistics*, 35, 765–805.
- Valdés, G. (2000). Introduction. In *Spanish for native speakers* (pp. 1–32). New York, NY: Harcourt College.

- Vasishth, S. & Nicenboim, B. (2016). Statistical Methods for Linguistic Research: Foundational Ideas - Part I. *Language and Linguistics Compass*, 10(8), 349–369.
- Wertz, C. A. (1977). *The Number of Genders in Polish*. Technical Report 1.
- White, L. (1989). *Universal Grammar and Second Language Acquisition*. Language Acquisition and Language Disorders. Amsterdam: John Benjamins Publishing Company.
- Willim, E. (2012). *Concord in Polish coordinate NPs as Agree*. Technical report.
- Yeni-Komshian, G. H., Flege, J. E., & Liu, S. (2000). Pronunciation proficiency in the first and second languages of Korean–English bilinguals. *Bilingualism: Language and Cognition*, 3(2), 131–149.
- Zagorska-Brooks, M. & Nalibow, K. (1970). The gender of referentials in Polish. *International Journal of Slavic Linguistics and Poetics*, 13, 136–142.
- Zombolou, K. (2011). Attrition in Greek Diaspora. In I. Ithemere (Ed.), *Language Contact and Language Shift: Grammatical and Sociolinguistic Perspectives*. Lincom Europa.