TESE DE DOUTORAMENTO

REPRESENTATION AND PROCESSING OF GRAMMATICAL GENDER: ANALYSING THE GENDER CONGRUENCY EFFECT

Ana Rita Sa Leite Dias

ESCOLA DE DOUTORAMENTO INTERNACIONAL DA UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

PROGRAMA DE DOUTORAMENTO EN PROCESOS PSICOLÓXICOS E COMPORTAMENTO SOCIAL

SANTIAGO DE COMPOSTELA

2021

DECLARACIÓN DA AUTORA DA TESE

Dna. **Ana Rita Sa Leite Dias**

Título da tese: **REPRESENTATION AND PROCESSING OF GRAMMATICAL GENDER: ANALYSING THE GENDER CONGRUENCY EFFECT**

Presento a miña tese, seguindo o procedemento axeitado ao Regulamento, e declaro que:

1) A tese abarca os resultados da elaboración do meu traballo.
2) De ser o caso, na tese faise referencia ás colaboracións que tivo este traballo.
3) Confirmo que a tese non incorre en ningún tipo de plaxio doutros autores nin de traballos presentados por min para a obtención doutros títulos.
4) A tese é a versión definitiva presentada para a súa defensa e coincide a versión impresa coa presentada en formato electrónico

E comprométome a presentar o Compromiso Documental de Supervisión no caso de que o orixinal non estea na Escola.

Santiago de Compostela, **24 de Outubro de 2021**

Sinatura electrónica

AUTORIZACIÓN DA TITORA E DIRECTORA DA TESE

Representation and processing of grammatical gender: Analysing the gender congruency effect

Dna. Isabel Fraga Carou

INFORMA:

Que a presente tese, correspóndese co traballo realizado por Dna. Ana Rita Sa Leite Dias, baixo a miña titorización e dirección, e autorizo a súa presentación, considerando que reúne os requisitos esixidos no Regulamento de Estudos de Doutoramento da USC, e que como directora desta non incorre nas causas de abstención establecidas na Lei 40/2015.

De acordo co indicado no Regulamento de Estudos de Doutoramento, declara tamén que a presente tese de doutoramento é idónea para ser defendida na modalidade de Monográfica con reprodución de publicacións, nas que a participación da doutoranda foi decisiva para a súa elaboración e as publicacións se axustan ao Plan de Investigación.

Santiago de Compostela, 24 de outubro de 2021

Sinatura electrónica

AUTORIZACIÓN DA DIRECTORA DA TESE

Representation and processing of grammatical gender: Analysing the gender congruency effect

Dna. Montserrat Comesaña Vila

INFORMA:

Que a presente tese, correspóndese co traballo realizado por Dna. Ana Rita Sa Leite Dias, baixo a miña dirección, e autorizo a súa presentación, considerando que reúne os requisitos esixidos no Regulamento de Estudos de Doutoramento da USC, e que como directora desta non incorre nas causas de abstención establecidas na Lei 40/2015.

De acordo co indicado no Regulamento de Estudos de Doutoramento, declara tamén que a presente tese de doutoramento é idónea para ser defendida na modalidade de Monográfica con reprodución de publicacións, nas que a participación da doutoranda foi decisiva para a súa elaboración e as publicacións se axustan ao Plan de Investigación.

Santiago de Compostela, 24 de outubro de 2021.

Sinatura electrónica

ABSTRACT

Grammatical gender is one of the most controversial issues in the study of language, having frustrated the work of many philosophers, linguists, and more recently, psycholinguists. Indeed, in recent decades special attention has been devoted to the link between grammatical gender and cognition. Yet, although the picture seems rather clear in terms of gender acquisition and retrieval during language comprehension, in the area of language production we are close to what we might call scientific havoc. In short, many of the on-going debates here involve disputes about the mandatory character of an agreement context for the retrieval of a noun's gender value, the automatic or competitive-based nature of the mechanisms behind gender selection, and different views on the link between morphophonology and gender processing. These matters have traditionally been assessed through the picture-word interference (PWI) paradigm, in which native speakers of different languages have to name aloud pictures using *target* nouns while ignoring superimposed written *distractor* words, and, more recently, through naming and forward translation tasks with learners of second languages (L2s). In all these tasks, conditions of gender congruency and incongruency are created through the matching or mismatching of the gender values of either targets and distractors or the equivalent translations of the L2 and the first language. Variations on the response times of the participants are expected depending on the congruency/incongruency between gender values.

More specifically, authors were looking for a gender congruency effect (GCE), by which pairs of nouns of the same gender speed up the response. However, as regards native language production, the presence and type of gender effects seems to depend on the language family and the type of response given by participants. In the case of Germanic and Slavic languages, the GCE is only observed with noun phrases that include elements of agreement, such as the singular definite articles in German ("der" [masculine] vs. "die" [feminine]). Critically, Caramazza et al. (2001) argue that because the so-called GCE depends on the presence of determiners, the effect reflects competition for selection between determiner forms of different

gender (a determiner congruency effect), rather than between gender values themselves. Hence, they conclude, gender values do not compete for selection and are retrieved automatically. For Romance languages, results are mainly null, except for a gender incongruency effect that is sometimes observed in Spanish and Italian when bare nouns are used to name the pictures (gender incongruency effect, GIE, facilitation for nouns of different gender). In bilingualism, the scenario is different, since a cross-linguistic GCE is consistently obtained regardless of the language family and the type of naming response involved (noun phrases vs. bare nouns). Hence, this reflects genuine effects of gender competition.

The main aim of this thesis is to shed light on this intricate group of findings and to discover the nature of the mechanisms underlying gender selection. To do so, we defined our object of study from both linguistic and psycholinguistic perspectives. We then proposed a clear definition of gender transparency based on gender phonology and morphology, as well as a standardized method that seeks to operationalize the degree of gender transparency of languages through an adaptation of Audring's (2019) concept of relative complexity of gender systems. We then conducted two meta-analyses, of the PWI paradigm and of naming/translation tasks, respectively. We did so by comparing the effect sizes related to the gender congruent and incongruent conditions through the computation of the Hedges' *g* measure. Results showed that the determiner congruency effect is of medium size and is highly significant, whereas the cross-linguistic GCE is small but also significant. Interestingly, the GIE was the smallest effect. We then conducted three studies with speakers of a hitherto unexplored language in this context, European Portuguese, an interesting choice in that it is a member of the Romance family and exhibits the same transparent gender cues as Spanish and Italian, but has one-to-one gender-determiner mapping, as in Germanic languages. Our first study included two experiments featuring PWI tasks to test the presence of gender effects with bare nouns, as well as the role of the animacy of the targets. A significant GCE, although one which was not supported by Bayesian analysis, was restricted to the condition featuring only inanimate nouns. Our second study included three PWI tasks that attempted to replicate the GCE obtained in Study 1, and also to test the determiner congruency effect with noun phrases, taking into consideration the role of ortho-phonological gender transparency. We failed to replicate the GCE but obtained a determiner congruency effect that interacted with the gender transparency of both targets and distractors. Finally, our third study consisted of a forward and backward translation task of bare nouns with early bilinguals of Portuguese and German. A cross-linguistic GCE was obtained

USC UNIVERSIDADE DE SANTIAGO

that depended on the proficiency of the two languages. Results were interpreted through the classical models of language production, the WEAVER++ (Levelt et al., 1999) and the Independent Network models (Caramazza, 1997), as well as through the Double-Selection Model (Cubelli et al., 2005), the Late Selection Hypothesis (Miozzo & Caramazza, 1999), and our own adaptation to language production of the Dual-Route Model (Gollan & Frost, 2001).

As a whole, the body of results can be seen to be better in line with the notion that gender values do compete. The case is clear when it comes to bilinguals, for whom the tasks used are quite simple, and for whom there is an increased flow of activation reaching gender nodes from multiple lexical entries. However, the answer is not so clear in the case of monolinguals: we detected notable limitations on the PWI paradigm, the complexity of which may render it incapable of capturing certain small effects of competition, such as those concerning gender values. Indeed, we note that (1) there is evidence that using more fine-grained techniques such as fMRI can result in competition between gender values being detected, even in German (see Heim et al., 2009); (2) there is an urgent need for the replication of findings in the field with new studies using larger participant samples, better stimuli control, and more precise measuring techniques; (3) the absence of effects of the PWI paradigm cannot be interpreted as evidence for the absence of effects of competition between gender values (Vasishth & Gerlman, 2021); thus (4) it might ultimately be the case that the field even requires a paradigm shift in terms of the experimental approach taken.

Keywords: grammatical gender, gender transparency, gender congruency effect, picture-word interefence paradigm, language production.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

FINANCIAL SUPPORT

This PhD was financed by the *Universidade de Santiago de Compostela* and the *Ministerio de Educación y Formación Profesional* through the grants "Ayudas para la Formación de Profesorado Universitario" (FPU16/06983), which financed these four years of scientific work, and "Ayudas complementarias de movilidad destinadas a beneficiarios del programa de Formación del Profesorado Universitario" (EST21/00165), which economically supported my research stay at the University of Radboud. My research, publications, and attendance to scientific events were funded by the group *Procesos Cognitivos y Conducta* through the grants given by the *Ministerio de Economía y Competitividad* (PSI2015-65116-P), the *Ministerio de Ciencia e Innovación* (PID2019-110583GB-I00) and the *Xunta de Galicia* (GRC 2015/006; ED431B 2019/2020). My research stay at the School of Psychology of the *Universidade do Minho* had zero costs thanks to the interuniversitary agreement between the PhD programs in basic Psychology held by the *Universidade de Santiago de Compostela* and the *Universidade do Minho* ("Acordo de cooperação entre a USC e a UMinho para mobilidade de professores e estudantes no âmbito do programa de doutoramento Processos Psicológicos e Comportamento Social da USC e do Programa de doutoramento Psicologia Básica da UMinho" — signed on the 10th of January of 2017). The *Centro de Investigação em Psicologia* (CIPsi) of the School of Psychology of the *Universidade do Minho* and the Donders Centre for Cognition of the Radboud University also supported part of the costs of all my experiments. Other costs related with experimentation and data analysis were supported by the Portuguese *Fundação para a Ciência e Tecnologia* (IF/00784/2013/CP1158/CT0013; UIDB/00305/2020) as well as the Portuguese Ministry of Science, Technology and Higher Education through Portuguese national funds and co-financed by FEDER through COMPETE2020 under the PT2020 Partnership Agreement (POCI-01-0145- FEDER-007653).

USC
DE SANTIAGO
DE COMPOSTELA

ACKNOWLEDGEMENTS

I was only six when my father told me to take a second look at the thousands of little stones standing at the shore of our usual beach. What at first seemed completely irrelevant and pointless to me, soon made me feverish with joy, curiosity, and excitement. They were of multiple colours, shapes, and textures. I became completely hooked searching for little stones and sharing them with my father: "*Look dad, this one is blueish and so round!*" I would say, and he would mirror my enthusiasm and look for an answer with me as to why that stone was how it was. My mom soon joined us, came up with a way to put some of these stones in a nice display in my room, and bought me books about geology. Nobody understood why I was so proud of these stones, why a kid would spend so much time just inspecting and asking questions about rocks, and later about trees, then about bugs, about mammals, and about people. My mom taught me to look at people the same way I looked at those beautiful stones. She was the one who explained to me that the boy constantly drawing attention on himself was an orphan; that the other kid who shoved me down probably saw his father doing the same thing to his mom; that the little girl who kept sneakily trying to copy my homework had learning disabilities and was ashamed of asking for help. Both of my parents taught me that everyone was different, everyone had a reason to be different, and that we should make an effort to try to understand, respect, and love everyone. My father, in his infinite patience, and my mother, in her infinite kindness, were not alone in the task of guiding me through life. Where my parents tried their best to control my overflowing ambition, for my brother it was never enough — I guess he is partially the reason my relation with achievement is both a problem and a blessing. During my childhood, he pushed me to my limits in many areas of life while continuously supporting me, he taught me not to ever be completely satisfied with my accomplishments, and then lent me a hand and waited for me every time I fell down so he could still guide me to a better version of myself.

For these and many other reasons, I want to start by thanking my family. Thanks to my parents, for making this possible, loving me unconditionally, feeding my curiosity, and fueling

USC
DE COMPOSTELA

what was already inside of me: an absolute thirst for knowledge. I believe this thirst is the basis of whatever kept me trying for years to understand a feature that has apparently driven so many people out of their minds and made many others give up. I have not and I will not give up, and for that, I want to especially thank my brother, for not making quitting an option. I also want to thank the three of them, Marcelo, Olinda, and Filipe for ensuring that I maintained my mother tongue while living abroad. This certainly played a part in turning me into the multilingual girl I am today trying her best to understand language.

It is undeniable that life presented me with a wonderful family. Yet none of this would have been possible if Camilo and his mom, Dolores, had not been there. To them I am completely and forever grateful. Camilo, there are simply no words to describe what we are to each other. With you, it is always fun and a little bit less of a struggle. Let us keep it that way for as much time as possible.

All those people tried their best to support me during this journey. Still, only those that go through a doctorate can truly understand what coping with life means as a young researcher. I was lucky to have the best companion one could ask for in this PhD journey: Doctor Cocas. We have a kind of *macho-frog* relationship going on so I cannot share my feelings here. I shall only say "*u rock, bro*". We will keep laughing at life together. I also want to thank all the lovely members of the Cognitive Processes and Behaviour Group, as well as my dearest colleagues from the Human Cognition Lab at the University of Minho. Special thanks to Aimé (a kind soul always willing to help me), Elena Andrade (another kind-hearted soul watching over my academic steps), and Helena Oliveira (for her continuous support during the first two years of this doctorate).

I confess that I cannot understand research as nothing other than the multi-disciplinary effort of many professionals working together to achieve the same goal. For this reason, my research would have not been possible without my colleagues and collaborators: Juan Haro, Juan Hernández-Cabrera, Carina Eira, Cristina Flores, Karlos Luna, Carlos Pinto, and Ângela Tomaz. To all of them, thanks for working with me, and coping with my multiple emails, calls for help, countless questions, and, sometimes, messy dramas. I would also like to thank Diego Pinal and Ana Paula Soares, as well as Kristin Lemhöfer for the supervision of my research stays at the University of Minho and the Radboud University, respectively. Diego, thanks for checking on me when I needed it the most. Kristin, you are the best supervisor one could ask for when doing a research stay abroad, especially during what was probably the lowest moment

USC UNIVERSIDADE DE SANTIAGO

of my PhD journey. At that time, I was living in a pandemic world and dealing with medical issues. Except for the day you looked like a "*broken egg*", you were a bright light on my path, and were able to colour my days, inspire me and teach me so much in so little time. I hope we meet again.

I have purposely saved for last the two people that held my hand tight in this journey. They are opposite sides of the same coin, yet they managed to keep me in balance through all these years.

To my *academic mother*, my supervisor Isabel Fraga: Thanks. Thanks for believing in me since the very beginning, for doing what was in your reach to bring me to this point. You have seen me flourish, and you made sure I had everything I needed to keep growing. Thanks for your endless patience with all my moments of anger and sorrow, and for your invaluable advice. I can only hope to be a researcher with at least a bit of your honesty and capability. You are a precious part of the cause behind all the opportunities I have had and will have in my life. From the moment you took me to the conference at the University of Minho when I had not even finished my degree, to this year, in which you still cared for me during your toughest times. In a way, you put all your faith in me, and refused to let me walk alone. Even at a distance, I assure you we will still be walking together. I shall finish with an advice of my own: Stop saying you are old! I demand many more moments of uncontrollable laughter by your side in the coming years, which I am sure, will be many.

To my *academic older sister*, my supervisor Montserrat Comesaña, I cannot properly express all the gratitude I feel for you taking me in with open arms when I needed it the most. Thanks for your innumerable teachings and for being a part of the reason I have flourished so much. Although sometimes frustrating, thank you for intellectually challenging me, constantly arguing with me, and tolerating all my yelling and breakdowns when your multiple counterarguments got on my nerves. You constantly pushed me forward, ensured that I did not waste any important professional opportunity, and made me laugh in every tough situation. I also want to thank you for giving me an unexpected new name, "Ruta", and for all the adventures together — at your side, these are always one of a kind. I believe we still have many other adventures to go through, the three of us. Montse, for you, I have a request: please, keep seeing the world with the kind eyes that have always characterized you.

Finally, if there is something I want to highlight here to whoever happens to be reading this is the following idea: life can be hard as hell, and no matter how capable, no personal

USC UNIVERSIDADE DE SANTIAGO

achievement would be possible if it weren't for the help of many others. Be humble enough to understand this, treat others nicely (you never know what they may be going through), help them, and be brave enough to accept the help of those willing to help you. I started my academic career at the age of 18; I left one degree to join another; I struggled financially and had classmates who periodically bought me groceries and covered many of my basic necessities when neither my family nor I could afford them; I studied and worked at the same time to support my parents, and had no time left for myself during the youngest years of my adulthood — yet I was gifted with good friends that made sure that every little moment I did have was filled with absolute happiness; I had the helping hands of professors that somehow cared for me and encouraged me on my journey, sometimes just by listening to me; I was lucky to live in a country whose government could aid me economically and could give me a chance of employment at the end of my degree, a chance that I took and that led me to this PhD. This PhD was far from easy, but the story repeated itself: I surrounded myself with marvelous people doing what they could to keep me going forward. My effort was immense, but the help I received was immense as well. For this reason: ***Obrigada* to everyone who made this possible**.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

TABLE OF CONTENTS

List of Figures ..... **XIII**

List of Tables ..... **XV**

List of Abbreviations ..... **XVIII**

1. Introduction **1**

2. Gender: Concept and characteristics **5**

2.1 Gender assignment 8

2.2 Gender agreement 11

2.3 Grammatical and natural gender: Introducing the perspective of Psycholinguistics 15

3. Gender Transparency **21**

3.1 Phonological gender transparency 21

3.1.1 The distinct role of the transparent cues "-a" and "-o" in gender acquisition and processing 23

3.1.2 Identifying cues for gender in other languages: The case of German and French 29

3.1.2.1 The case of German 30

3.1.2.2 The case of French 33

3.1.3 Conclusions on phonological gender transparency 39

3.2 Matching of gender phonology and morphology 40

3.3 Standardizing gender transparency 43

3.3.1 Factors to assess the degree of gender transparency of a language 44

3.3.2 Definition of gender transparency and summary of the determining factors 48

4. Models of representation of grammatical gender **51**

4.1 Word-form Encoding by Activation and Verification Model (WEAVER++) 51

4.2 Independent Network Model 59

5. Exploring grammatical gender in language production **63**

5.1 Evidence on native language production 64

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

5.1.1 Studies with Germanic and Slavic languages 68
5.1.2 Studies with Romance languages 80
5.1.3 Discussion on native language production 89
5.1.4 Meta-analysis of the monolingual data 97
5.1.4.1 Method 98
5.1.4.1.1 Literature search 98
5.1.4.1.2 Meta-analytic approach 105
5.1.4.2 Results 107
5.1.4.2.1 Main meta-analysis: The GCE 107
5.1.4.2.2 Meta-Analysis on the singular-plural paradigm 108
5.1.4.3 Discussion 109
5.2 Evidence on second language production 113
5.2.1 Meta-analysis of the bilingual data 123
5.2.1.1 Method 124
5.2.1.1.1 Literature search 124
5.2.1.1.2 Meta-analytic approach 126
5.2.1.2 Results 128
5.2.1.3 Discussion 129
5.3 Summary of the monolingual and bilingual data 132
6. Grammatical gender in European Portuguese 135
6.1 The grammatical gender system of European Portuguese 138
6.2 Study I: Bare nouns and animacy 161
6.2.1 The role of animacy in the picture-word interference (PWI) paradigm and gender processing 163
6.2.2 The present study 166
6.2.3 Experiment 1 168
6.2.3.1 Method 168
6.2.3.1.1 Participants 168
6.2.3.1.2 Materials 168
6.2.3.1.3 Procedure 173
6.2.3.2 Results 174

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

6.2.4 Experiment 2 177
6.2.4.1 Method 177
6.2.4.1.1 Participants 177
6.2.4.1.2 Materials 177
6.2.4.1.3 Procedure 177
6.2.4.2 Results 180
6.2.3 Discussion 182
6.3 Study II: Noun phrases and bare nouns plus gender transparency 188
6.3.1 Experiment 1 195
6.3.1.1 Method 195
6.3.1.1.1 Participants 195
6.3.1.1.2 Materials 195
6.3.1.1.3 Procedure 200
6.3.1.2 Results and interim discussion 201
6.3.2 Experiment 2 205
6.3.2.1 Method 205
6.3.2.1.1 Participants 205
6.3.2.1.2 Materials 205
6.3.2.1.3 Procedure 205
6.3.2.2 Results and interim discussion 205
6.3.3 Experiment 3 208
6.3.3.1 Method 208
6.3.3.1.1 Participants 208
6.3.3.1.2 Materials 208
6.3.3.1.3 Procedure 212
6.3.3.2 Results and interim discussion 212
6.3.4 Discussion 213
6.4 Study III: Cross-linguistic GCE and the impact of proficiency 217
6.4.1 Method 220
6.4.1.1 Participants 220

USC DE COMPOSTELA

6.4.1.2 Materials 220
6.4.1.3 Procedure 224
6.4.2 Results 225
6.4.3 Discussion 227
7. General discussion **229**
7.1 Hypotheses for the mechanisms underlying gender selection 233
7.1.1 Gender nodes do not compete for selection 233
7.1.2 Gender nodes compete for selection 237
7.2 Controversial aspects of the literature 242
8. Concluding Remarks **245**
8.1 Limitations of this thesis 245
8.2 Ideas for future research 246
8.3 We are on the right path 248
References **251**
Appendix A: Experimental materials Study I **289**
Appendix B: Experimental materials Study 2 .....**298**
Appendix C: Experimental materials Study 3 .....**303**
Appendix D: Information on published papers (*relación de publicacións*) **307**
Appendix E: Permission to use and reproduce published articles **309**
Resumo en galego **312**

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

LIST OF FIGURES

Figure 1. Gender values in Dyirbal 6
Figure 2. The Animacy Hierarchy 19
Figure 3. Proposal of a continuum of gender transparency 50
Figure 4. Simplified structure of lexical access according to WEAVER++ 53
Figure 5. Representation of the two self-monitoring routes of lexical access in WEAVER++ 56
Figure 6. Simplified structure of lexical access according to the IN model 60
Figure 7. Stages of lexical access in the PWI paradigm as conceived by the WEAVER++ model 65
Figure 8. Lexical access according to the Double-Selection Model 85
Figure 9. Primed Unitized Activation Hypothesis 92
Figure 10. Funnel Plots for Studies with Germanic/Slavic Languages (Panel A) and with Romance Languages (Panel B) 108
Figure 11. Representation of the gender-integrated hypothesis for words of the same gender across languages (panel A) and those of different gender (panel B) 114
Figure 12. Representation of the gender-autonomous hypothesis for words of the same gender across languages (panel A) and those of different gender (panel B) 114
Figure 13. Forest plot with the effect sizes of GC on RTs for all studies with the meta-analytic estimate 128
Figure 14. Representation of the routes of gender retrieval within the IN model 192
Figure 15. Representation of the routes of gender retrieval within WEAVER++ 192
Figure 16. How L2 gender representation develops during acquisition following the BIA-d model 219

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Figure 17. Interaction between gender congruency and difference of proficiency between languages.....227

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

LIST OF TABLES

Table 1. Summary of the differences between grammatical and natural gender 20
Table 2. Morphological examples of the Spanish natural gender and agreement system 41
Table 3. Factors to operationalize the degree of gender transparency of languages 48
Table 4. Summary of the main characteristics of the WEAVER++ and IN models 62
Table 5. Summary of all reviewed studies on the GCE for the main meta-analysis 101
Table 6. Summary of all the reviewed studies on plural gender-gain effect for the additional meta-analysis 104
Table 7. Description of the experiments conducted on the GC effect with bilingual population 119
Table 8. Distribution of the nominal endings in European Portuguese according to orthography 141
Table 9. Distribution of the nominal endings in European Portuguese according to phonology. 145
Table 10. Distribution of the nominal endings according to orthography for the most frequent nouns of European Portuguese 150
Table 11. Distribution of the nominal endings according to phonology for the most frequent nouns of European Portuguese 152
Table 12. List of European Portuguese derivational suffixes according to gender 155
Table 13. Operationalization of the degree of gender transparency of European Portuguese 159
Table 14. Conditions in each of the experiments 167
Table 15. Mean and standard deviations (in parenthesis) of the controlled variables per target type for all the experimental pictures in Experiment 1 170

USC UNIVERSIDADE DE SANTIAGO

Table 16. Mean and standard deviations (in parenthesis) of the controlled variables for the distractors across conditions of Target Gender and Gender Congruency within every condition of Animacy Presence in Experiment 1.....172

Table 17. Mean RTs and standard errors (SE) for each condition in Experiment 1.....175

Table 18. Mean and standard deviations (in parenthesis) of the controlled variables across targets' gender values for all the experimental pictures in Experiment 2.....178

Table 19. Mean and standard deviations (in parenthesis) of the controlled variables for the distractors across conditions of Target Gender and Gender Congruency in Experiment 2....179

Table 20. Mean RTs and standard errors (SE) for each condition in Experiment 2.....180

Table 21. Mean RTs and standard errors for animate and inanimate nouns in Experiment 1..... 185

Table 22. Most likely outcomes of the experiments regarding the factor of gender congruency. 194

Table 23. Means and standard errors of the controlled variables per target type for the 64 experimental pictures in Experiment 1 196

Table 24a. Means and standard errors of the controlled variables for distractors per experimental condition of feminine targets in Experiment 1.....198

Table 24b. Means and standard errors of the controlled variables for distractors per experimental condition of masculine targets in Experiments 1 and 2 199

Table 25. Mean response latencies (in milliseconds, standard deviations in parentheses) for Experiment 1 (noun phrases) 202

Table 26. Mean response latencies (in milliseconds, standard deviations in parentheses) for Experiment 2 (bare nouns).....206

Table 27. Means and standard errors (SE) of the controlled variables per target type for the 32 experimental stimuli in Experiment 3.....209

Table 28. Means and standard errors (SE) of the controlled variables for distractors per experimental condition of masculine targets in Experiment 3.....211

Table 29. Mean response latencies (in milliseconds, standard deviations in parentheses) for Experiment 3 (bare nouns and opaque targets).....212

Table 30. Means and standard errors of the controlled variables for Portuguese nouns across translation types 221

USC DE SANTIAGO DE COMPOSTELA

Table 31. Means and standard errors of the controlled variables for German nouns across translation types 223
Table 32. Means and standard errors of the overlap measures 223
Table 33. Mean RTs and SEs of gender congruency according to target language and target gender 225
Table 34. Estimated fixed effects for the final model 226

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

LIST OF ABBREVIATIONS

AOA = Age of acquisition

DSIH = Determiner Selection Interference Hypothesis

DSM = Double-Selection Model

GCE = Gender congruency effect

GIE = Gender incongruency effect

EEG = Electroencephalography

F = Feminine

fMRI = Functional magnetic resonance imaging

GSIH = Gender Selection Interference Hypothesis

IN = Independent Network model

L1 = First language

L2 = Second language

LSH = Late Selection Hypothesis

M = Masculine

Ms = Miliseconds

N = Neuter

O = Opaque

PWI = Picture-word interference paradigm

RTs = Reaction times

T = Transparent

SOA = Stimulus Onset Asynchrony

WEAVER++ = Word-form Encoding by Activation and VERification model

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

1 INTRODUCTION

From ancient Greek times, through the Middle Ages, and up to the present day, grammatical gender has been the subject of highly polarized debate. It is unclear how and why this feature emerged in an analogous way in so many of the languages of the world. Its function and usage in language are still relatively little understood. Its description, which came to be of greater interest to linguistics in the 20th century, has in turn led to many disputes regarding the question of how it differs from other grammatical features. More recently, the link between grammatical gender and cognition has fallen under the gaze of psycholinguists, this prompting a new focus in heated discussions on the representation and processing of grammatical gender in the human mind during both language comprehension and production.

According to Aristotle, Protagoras was the first to give a name to what speakers were doing when treating nouns differently in terms of agreement (Aristotle, Rhetoric, 1407b/1984). Protagoras was a sophist Greek philosopher who, among many other contributions to grammar, made a distinction between words that he labelled masculine, feminine, and neuter. Little did he know that he was thus identifying a phenomenon which would not only become one of the most controversial in language studies, but which would also bear these terms for millennia. From the moment that Protagoras noted the existence of gender, it is difficult to find any generous words devoted to the issue from the many things that have been said about it. Kilarski (2005) provides a throughout overview of all types of ill-favoured accusations here, showing that the terms used to refer to grammatical gender share a common ground of frustration. It has been described as an unessential and meaningless category with no useful purpose (Ibrahim, 1973; of Ockham, n.d./1980), an unnecessary component of language best to be dispose of (Adelung, 1783), a luxury (Bally, 1935; von der Gabelentz, 1891), a superfluous feature (Fodor, 1959), and even a historical accident, a deformity, a bad habit, the cause for many illnesses, including nightmares, pathological behaviours, erotic and religious delusions, sadism, and savagery (de Courtenay, 1929; Kilarski, 2005, p. 29). This type of quasi-personal approach to grammatical gender can still be seen in modern times: Trudgill, in a moment of enlightenment,

USC UNIVERSIDADE DE SANTIAGO

described grammatical gender as “linguistic male nipples” (1999; p. 149) and Boutonnet and colleagues (2012) concluded, somewhat hopelessly, that gender escapes logic.

Indeed, the fact that the word “puzzling” has become the ideal linguistic crutch when introducing grammatical gender to an audience should not surprise us (Audring, 2009, 2014; Corbett, 1991; Curzan, 2003; Kramer, 2015; Mills, 1986; Siemund, 2013; Trudgill, 1999). It is not entirely clear though, whether we are building a puzzle or rather seeking to solve it, this as a means of understanding how the whole thing evolved in the first place. Certainly, it might appear discouraging to write a doctoral thesis on a matter that has remained highly controversial since long before the Christian era. I have often found myself wondering what our contribution could possibly be to the study of a phenomenon that has inspired such an array of negative sentiments from so many philosophers and researchers. In fact, I will neither confirm nor deny that after four years of an intense relationship with grammatical gender, I may have come to an empathetic understanding of most of the claims mentioned above. However, in what follows, I will make every effort to keep the ugliness of grammatical gender to myself, while expressing all the pleasure and excitement that the study of this feature has afforded me during my research.

This thesis takes a psycholinguistic perspective on the assessment of grammatical gender within the modality of language production. More specifically, it studies the representation and processing of grammatical gender during the lexical access of nouns in oral production. We hereby intend to shed some light on these challenging issues:

- The role of agreement contexts in the retrieval of gender values during the production of nouns.
- The role of phonological gender regularities found in nominal endings on the processing of grammatical gender during lexical access in language production.
- The reasons behind the cross-linguistic differences that seem to exist when exploring the above issues (i.e., the role of agreement and gender form-based regularities).
- The role of the general degree of gender transparency of each language in the acquisition, representation, and processing of grammatical gender.

To this end, the thesis focuses on the so-called gender congruency effect, obtained through the classic picture-word interference (PWI) paradigm, as well as on the cross-linguistic gender congruency effect, derived through simple naming tasks and translation tasks for bilinguals. These effects and tasks will be described subsequently. We present several theoretical reviews and meta-analyses, considering both native speakers and learners of second languages (L2s).

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

We also provide the results of a series of PWI and translation tasks conducted with native speakers of European Portuguese, filling a void in the research literature for this language. Finally, we draw conclusions, taking into consideration both the results of the meta-analyses and the results of the experiments conducted in European Portuguese.

The thesis is organized as follows:

Chapter 2 opens with a detailed definition and typological assessment of grammatical gender from a linguistic perspective, drawing the reader's attention to the two main dimensions of this feature: assignment (subchapter 2.1) and agreement (subchapter 2.2). Distinctions are established between gender and two other main grammatical features to which it is closely related: classifiers and inflective classes. We then present a psycholinguistic perspective on gender, highlighting the opposition between grammatical and natural gender (subchapter 2.3) and detailing their different cognitive repercussions. The chapter ends by setting out the notion of grammatical gender as the main object of study in this thesis.

Chapter 3 addresses one of the most significant characteristics of grammatical gender that diverges across gendered languages: gender transparency. We propose a standardized definition of this concept based on the relation between gender and phonology (subchapter 3.1), as well as morphology (subchapter 3.2), taking into account the impact that this relation may have on the acquisition and processing of gender. We end by presenting a theoretical proposal based on the concept of relative complexity to operationalize the degree of gender transparency across languages.

Having comprehensively defined the concept of gender and gender transparency from the perspective of linguistics and psycholinguistics, Chapter 4 presents the architecture and functioning of lexical access according to the cognitive models of language production that have considered gender within their tenets. We thus consider the WEAVER++ model (subchapter 4.1), and the Independent Network model (subchapter 4.2), giving special attention to their conceptions of the representation and processing of grammatical gender.

Chapter 5 presents the experimental assessment of grammatical gender as a lexical component of nouns during language production. To do so, we explore the literature that tests the classical gender congruency effect (henceforth, GCE) obtained with monolinguals through the PWI paradigm (subchapter 5.1), as well as the literature testing the cross-linguistic GCE obtained with bilinguals through naming and forward translation tasks (subchapter 5.2). Two meta-analyses are included within each subchapter. Results are interpreted in relation to the

USC UNIVERSIDADE DE SANTIAGO

models presented in Chapter 4 We draw several conclusions that allow us to pinpoint some specific, unanswered research questions (subchapter 5.3).

In Chapter 6, these research questions are explored through a series of production tasks featuring the PWI paradigm and testing native speakers of European Portuguese, as well as a translation task with early bilinguals of European Portuguese and German. We start by presenting a systematical assessment of the characteristics of the gender system of European Portuguese (subchapter 6.1). We then show the results of two experiments testing gender processing and the role of animacy within the PWI paradigm itself (subchapter 6.2) and three experiments testing the role of agreement and phonological gender transparency in gender processing within the same paradigm (subchapter 6.3). Finally, we assess the impact of proficiency in both European Portuguese and German on the cross-linguistic GCE (subchapter 6.4).

In the General Discussion, we present the main conclusions regarding the representation and processing of grammatical gender during the lexical access of nouns in language production based on previous reviews and results (section 7.1). We then propose different hypotheses of grammatical gender processing in language production (section 7.2) and analyse all the loose or unresolved aspects of the literature that may be critical to opt for one hypothesis over another (section 7.3).

Finally we address what we believe are the limitations of this thesis (section 8.1) and propose a few ideas for future research to test the final hypotheses and to overcome the problematic aspects previously identified (section 8.2). We end by presenting a brief reflection on how we have firstly found and then left the state of the art, as well as by making prospective analysis of what will be the future of this field (section 8.3).

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

2 GENDER: CONCEPT AND CHARACTERISTICS

Among all the possible grammatical features of a language, gender stands as one of the most peculiar. Grammatical features are the elements into which "words can be broken down" (Corbett, 2012). These features commonly refer to the properties of words and are necessary for the construction of well-formed and comprehensible messages within a specific language. Grammatical features include nominal and verbal number, case, verbal tense, definiteness, person, animacy, and gender (see Kibort & Corbett, 2008). Whereas some features, such as number, are present in most languages (there are at least four languages without number, see Corbett, 2000), others, such as case, are not found so widely. Gender, however, seems to be present in approximately half of the languages of the world according to the most recent geographical study, based on a representative sample of 256 languages (Corbett, 2013a; see also Corbett, 2008). These languages are labelled as "gendered". Most gendered languages possess a gender system that classifies nouns according to two gender values, although less frequently we can find more[1]. Dual gender systems are the most common worldwide (Corbett, 2013a), especially within Afro-Asiatic (e.g., Hebrew) and Indo-European language families. The Indo-European family includes the highest number of speakers across the gendered families of the world (Corbett, 2013a). Within this group, in the Romance branch, languages such as Catalonian, French, Galician, Italian, Portuguese and Spanish all have two gender values: masculine and feminine. For instance, in Portuguese, nouns can be either masculine, such as "sumo" (juice) or "menino" (male kid) or feminine, such as "casa" (house) or "menina" (female kid). Similarly, there are Indo-European languages outside Europe with masculine and feminine gender values, such as Hindi and Punjabi (e.g., both juice and house in Punjabi are masculine, "ਜੂਸ" ["Jūsa"] and "ਘਰ" ["ghara"]). However, three-gender systems are also common, as it is

[1] There are languages that have two gender systems. One of these is the Austroasiatic language Santali, which has a gender system based on animacy, and another such system based on sex-distinctions, this due to language contact with Indo-Aryan (Ghosh, 2008; Sinnemäki, 2019).

the case of Czech, Icelandic, German and Russian, which have masculine, feminine and neuter values. For instance, in German, "juice" is masculine, but "house" is neuter ("der Saft" and "das Haus", respectively); an example of a German feminine noun would be "door" ("die Tür", the door), which is also feminine in Portuguese (porta) but masculine in Punjabi ("ਦਰਵਾਜ਼ਾ", "Daravāzā"). There are also some languages within the Germanic and Slavic branches (such as Danish, Dutch, and Swedish) in which the masculine and the feminine gender values have collapsed, resulting in a dual gender system consisting of common and neuter nouns.

In a representative sample of 112 languages from around the world, Corbett (2013a) finds that systems with two and three gender values are the most common (n = 50 and n = 26, respectively). However, there are gender systems that show a greater number of values, as can be seen in other families, such as the Austro-Asiatic, Austronesian, Caucasian (e.g., Nakh-Daghestanian family), Dravidian, Khoisan, Niger-Congo, Nilo-Saharan, and Papuan languages. For instance, Dyirbal is an Australian language with four gender values (see Figure 1). The Niger-Congo language Fula stands out here, with around 20 genders, depending on the dialect (Corbett, 1994; 2013b).

Figure 1

Gender values in Dyirbal

Gender I	Male humans; non-human animates
Gender II	Female humans; fire; water; battle-related terms
Gender III	Edible fruits and vegetables
Gender IV	Other nouns, usually inanimate (RESIDUAL CLASS)

Importantly, one major source of disagreement in the literature concerns the name given to the phenomena of gender in languages such as Fula and other Niger-Congo, Australian, Bantu, Caucasian, and Sudanic languages. Some authors consider that rather than gender values, we should call them noun classes (Hellinger and Bußmann, 2001; Ibrahim, 1973). Gender derives from the Latin "genus", which originally meant "kind" or "sort" (Corbett, 2009), and was initially assigned only to the Proto-Indo-European noun classification system, which distinguished between the common gender (later usually divided in masculine and feminine) and the neuter gender. Indo-European languages, as well as Dravidian, Afroasiatic, and some Native American languages, have systems that adapt to this more restricted definition of gender. However, many linguists prefer to use the term "noun class" when referring to languages that

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

classify nouns according to more than three values (e.g., Australian, Niger-Congo, Caucasian, Khoisan, and Papuan languages, among others). In any case, we can identify only two differences between one type of system and another: (1) when there are more than three values, these are named following a numerical sequence (see Figure 1 for an example in Dyirbal), and (2) the semantic divisions between these classes are richer than in the case of languages with only two or three values. However, we might note that in Dyirbal, for instance, each class has been given a name, further reducing the apparent differences between gender and noun classes: (I) bayi, (II) balan; (III) balam, (IV) bala (Dixon, 1972). Besides, recent investigations claim that some Indo-European languages may have had four genders after evolving from Latin (see Loporcaro & Paciaroni, 2011). Hence, the dissimilarities between gender and noun class seem to be mere nuances, and opting for one term or another entails a trivial matter of linguistic tradition, as Corbett has argued (2009; 2013a). This is especially evident when we verify that, from the point of view of the real repercussions for the structure and functioning of a language, gender and noun classes do not differ (Corbett, 2013a). More specifically, in all languages with this grammatical feature (either gender or noun class), there are two defining factors (Alvanoudi, 2014):

- Nouns are classified according to a number of categories following semantic and sometimes morpho-phonological principles.
- Other elements of the speech must adapt their form to agree with the gender value/noun class assigned to the head noun. Hence, in Portuguese: "**A** (F) casa (F) pequen**a** (F)" (The small house) but "**O** (M) copo (M) pequen**o** (M)" (The small glass).

Nowadays, it is common to refer to this grammatical feature as gender, and use "noun classes" as a synonym. There are some authors, though, who prefer to consider gender as a type of noun class with a small system of values, and hence languages with gender could be said to have noun classes whereas languages with noun classes could not be automatically said to have gender (Aikhenvald, 2000; Grinevald, 2000). The inverse is also possible, with some authors using the term "noun class" as a "subtype of gender systems" (Sinnemäki, 2019, p. 139). Because the definition of gender and noun classes is identical, and the repercussions for language production and comprehension do not differ, in this thesis we will follow the most accepted nomenclature and consider gender to be the same as noun classes (Corbett, 1991). Consequently, in all the abovementioned languages we will define gender as this grammatical characteristic of nouns expressed through two or several values, and that it is realized outside

USC UNIVERSIDADE DE SANTIAGO

the nouns themselves through a phenomenon called "concord" or "agreement" (Corbett, 1991; Hockett, 1958; Kramer, 2015; Nichols, 1992; Steele, 1978). This definition captures the two focal points of gender, which lead us to assess this feature from two different perspectives: that of the nouns themselves (gender assignment), and the perspective of the other words contained in speech (gender agreement).

2.1 GENDER ASSIGNMENT

Nouns can be assigned to gender values following either semantic or morpho-phonological rules. Gender systems whose assignment rules are purely semantic tend to rely on the split into male and female, or animate beings and inanimate objects (or both).[2] For instance, in Tamil, a Dravidian language spoken in the state of Tamil Nadu and the region of Puducherry in India, there is a gender system in which nouns are assigned to three different gender values following strictly semantic rules: masculine for male rational beings, feminine for female rational beings, and neuter for non-rational beings and inanimates. Other systems are predominantly, rather than strictly, semantic. As previously seen in Dyirbal, nouns are assigned to each gender value following semantic rules, but one of these values is what Corbett (1991) calls the "semantic residue", that is, it includes nouns that are not classified according to positive semantic criteria. Since these nouns do not follow any of the semantic criteria, they are assigned to an umbrella value with no fundamental semantic drive other than not fitting into the other values. Finally, there are gender systems that follow both semantic and morpho-phonological assignment rules. In fact, formal systems do not exist on their own (Corbett, 1994), which has led many linguists to argue that gender systems always have a semantic core (Aksenov, 1984; Audring, 2017; Corbett, 1991). One of the languages in which gender can be almost unanimously predicted on the basis of phonological rules is that of Qafar, an East Cushitic language spoken in Ethiopia and Djibouti, which has a dual gender system (Corbett, 1991). In Qafar male beings are linguistically masculine and female beings are feminine. The remaining nouns are assigned according to the following rule: Nouns that end in an accented vowel are feminine, such as amoytà (queen) and qaadà (tradition); others are masculine, such as amoyta (king) and minin (eyebrow). One might argue that phonological rules would be enough to establish assignment,

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

[2] Other semantic divisions are possible. As noted by Corbett (1991), De La Grasserie proposed eight main types: animate/inanimate, rational/non-rational, human/non-human, male human/other, strong/weak, augmentative/diminutive, male/other, masculine/feminine/non-sexed. These are all the existing combinations, yet it is possible to go further (large/small, edible, insect, etc.; Corbett, 1991).

but there are exceptions that do not allow this simplification to be wholly accurate. In these exceptions, semantic and phonological rules are in conflict, and the semantic ones always override the phonological criteria, thus a noun like “abbà” (father) is masculine (Corbett, 1991).

Other examples of gendered languages that combine semantic and morpho-phonological assignment rules are found in the Indo-European family. For example, in Russian, nouns can be either masculine, feminine, or neuter, and assignment is both semantically and morphologically driven. Nouns that refer to humans are classified according to the biological sex of the referent, so that “Babushka” (grandmother) is feminine and “Dedushka” (grandfather) is male. In the other cases, nouns are assigned to all the genders following morphological principles: nouns of declension 1 are masculine, nouns of declensions 2 and 3 are feminine, and nouns of declension 4 are neuter. Note that also in Russian, even though “Dedushka” belongs to the second declension (“-a”) the semantic rule is given priority, and since the noun refers to a male human, it is assigned the masculine value. In German, the morpho-phonological rules are not as straightforward as they seem in Russian. More specifically, nouns can be assigned to gender values following semantic, morphological, and phonological regularities (Köpcke, 1982; Köpcke & Zubin, 1983, 1984). The semantic assignment system is based on the biological sex (male and female). Importantly, note that languages in which the basic semantic rule is sex can display other semantic regularities for gender assignment. For example, in German, other semantic fields have been identified with specific gender values: colour names and metals are neuter (e.g., in Spanish and Portuguese, they are masculine), and trees and flowers tend to be feminine (the same follows in Spanish and Portuguese). However, the morpho-phonological assignment regularities are quite complex and usually associated with monosyllabic nouns. Mono-syllabic nouns tend to be masculine, especially when phonologically ending in [ft], [xt], [cˌt] and in [u:r] or [y:r]. However, nominal beginnings also seem to correlate with gender values, for example, words phonologically starting with [kn], [dr] or [tr], with [s]+consonant, and those ending on a nasal+consonant, as well as those with two initial and two final consonants, tend to be masculine (Szagun et al., 2007). Similarly, the orthographical nominal endings “-e” and “-ie” tend to be feminine, whereas the nominal endings “-en”, “-er” and “-el” are usually masculine, and nouns ending in a stressed vowel followed by /m/ or /l/ are neuter (for an extensive overview, see Corteen, 2019). Furthermore, in German there are derivational morphemes that also correlate with gender. These are affixes that are added to lexemes to create new words and hence change their

USC UNIVERSIDADE DE SANTIAGO

original meaning or function (word class). For example, as described in Szagun et al. (2007), nouns ending in "-chen" and "-lein" (to form diminutives) are neuter. These are only a few examples of some regularities that can be found in the German gender assignment system; however, there are many more regularities found in compounds, reduced nouns, and nominal inflections, among others. Additionally, plurals have their own form regularities and German has four cases, which alters the plural formation and may overlap regularities of different genders. For example, although "-en" is commonly a phonological cue for the masculine gender in the nominative case, it is also the marker for the dative plural form, and so nouns like "Geschenk" (gift), which is neuter, forms its plural in the dative as "Geschenk**en**" (gifts).

Because gender assignment can be based on formal regularities, gender is said to be more or less predictable depending on how visible the gender value of a noun is in its form. When gender values are evident in the form of nouns, then the gender system of a language is considered overt. When they are not, then it is said that the language has covert gender. Importantly, as Corbett (1991) stated, it is difficult to make a strict classification of a gender system as overt or covert. Instead, we can say that gender is more or less overt. For example, the Russian gender system has been described as moderately overt (Corbett, 1991), whereas that of Qafar is extremely overt. Yet the existence of many form regularities does not automatically make a gender system overt. For instance, although we have seen that German has many morpho-phonological rules that make gender directly correlated with the form of nouns, when it comes to locate it in a scale of overtness, it would fall rather on the covert side. Similarly to French, the vast quantity of regularities and rules make the phonological assignment of gender quite complex, with many exceptions and ambiguous endings, apart from the existence of cases. As Corbett (1991) notes, an ideal overt system would be one with a consistent and simple marker for gender, for example, all masculine nouns ending in "-o" and feminine ones ending in "-a". Indeed, within the Romance branch, Italian, Spanish, and Portuguese are quite overt in this sense. Still, even in these languages, there are other morpho-phonological regularities as well as ambiguous endings that prevent us from stating that the gender systems of these languages are purely overt. For instance, in Portuguese, Spanish, and Italian, nouns ending in "-e" can be either masculine or feminine, without great correlational differences between gender values. The morpheme "-aje" in Spanish transforms verbs into masculine nouns (e.g., land and landing: "aterrizar" → "aterrizaje" [M]), whereas in Portuguese, "-agem" creates feminine nouns (land and landing: "aterrar" → "aterragem" [F]).

USC UNIVERSIDADE DE SANTIAGO

The greater overtness of these languages, however, derives from the fact that the system of rules and regularities is simpler. The question of how the complexity of a gender system affects its overtness has already been addressed by many researchers, and we highlight here the work of Audring (2017) based on the principles of Canonical Typology (Corbett 2006, 2012; Corbett & Fedden, 2015). We will consider the relation between linguistic complexity and gender overtness in the second chapter of this thesis.

2.2 GENDER AGREEMENT

The most important characteristic of gender is agreement: the gender values assigned to nouns have different formal repercussions for a set of diverse elements contained in speech. These elements are called agreement targets, whereas the head nouns they depend on are called controllers (Corbett, 1991). Agreement targets may include adjectives, numbers, demonstratives, articles, numerals, possessives, relative and personal pronouns, adverbs, adpositions (prepositions and postpositions), completizers (e.g., conjunctions), and verbs (in any form, especially participles; Corbett, 1991). Note that the word classes agreeing in gender with the noun vary from one language to another. For instance, in the Slavic branch of the Indo-European family, verbs usually agree in gender with the subject. In Russian, the ending "-a" (among others) is usually added to the feminine form of the past tense due to agreement:

(1)

Zhenshchina spal**a**. "The woman slept"	Zhenshchina znal**a**. "The woman knew"
Muzhchina spal. "The man slept"	Muzhchina znal. "The man knew"

In the Romance and Germanic branches, however, this does not occur. For instance, in German, the first sentence of the Russian example above (1) would show agreement uniquely on the definite articles: "Die/der Frau/Mann schlief". Agreement is commonly realized through inflectional affixes attached to a stem as a preceding element (prefixes) or a proceeding element (suffixes), or both. The most common form of gender marking is through suffixes, as seen in the Russian and German examples. In Indo-European languages, this is the norm. For example, in Portuguese, masculine is marked with "-o" (or "empty" index - no marking) and feminine is marked with "-a" (e.g., "aquel**e** [M] menin**o** [M] alt**o** [M] foi **o** [M] últim**o** [M] em chegar" but "aquela [F] menin**a** [F] alt**a** [F] foi **a** [F] últim**a** [F] em chegar" [that tall female/male kid was the last to arrive]). On the other hand, prefixes can also be used. In Mufian, an Arapesh language from Papua New Guinea, 17 gender values have been identified, and verbs are marked with

USC DE COMPOSTELA

prefixes depending on the class (e.g., class 1 is marked with singular verbal prefix "l-" and class 2 with "g-"; Alungum & Lukas, 1978, p. 93). There is also a rarer form of agreement marking through infixes (inside the morphological structure of the word; for more detail, see Corbett, 1991, p. 116).

Indeed, although gender can be visible in the form of the noun itself (in the case of a sufficiently overt gender system), this is not a requirement of gender, but agreement is visible and should be the ultimate source for gender deduction (Audring, 2017). As Audring (2008) states, when the nominal properties (e.g., phonological regularities) and the agreement markers apparently suggest different genders, the elements of agreement must be taken as indicative. So, in the case of the Portuguese noun phrase "a tribo" ("the tribe"), although the nominal ending "-o" is a phonological regularity for the assignment of the masculine value, since the definite article is in its feminine form ("a", and not "o"), the gender of the noun is feminine.

Importantly, agreement constitutes the grounds for the distinctiveness of gender when it comes to differentiating it from other categorization systems. On the one hand, if nouns can be categorized according to different semantic-based classes but agreement is completely absent, then we are not looking at gender but another type of "classifier" (Aikhenvald, 2000; Grinevald, 2000). Gender, besides being a grammatical feature, belongs to a group of devices called "classifiers" that categorize nouns and nominals (Aikhenvald, 2000, 2019). Apart from gender, these include numerical, possessive, relational, verbal, deictic, and noun classifiers. They classify nouns following mostly a semantic principle (Aikhenvald, 2000; Corbett, 2013a). Numerical, possessive, verbal and noun classifiers are particularly frequent (Seifart, 2010, p. 721). For instance, numerical classifiers can be found within noun phrases adjacent to the numeral itself, usually to quantify either uncountable nouns that cannot occur directly with numerals or even countable nouns (Greenberg, 1977; Fedden & Corbett, 2017). When we are speaking English and we wish to allude to a specific number of dolls, we can say "one doll" or "two dolls", but in the case of uncountable nouns this does not apply, and instead of "one sand" and "two sands" we use a measuring element such as "one or two pounds of sand". Some languages have developed numerical classifiers, an extra grammatical element that is added to the noun phrase to quantify these uncountable cases and which has even been generalized to countable nouns like "doll". For instance, as noted by Kung (2007) for Lhiimaqalhqama',

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

numeral classifiers are prefixed to the numerals, and speakers use them as follows:

(2)

“one sheet of paper” as “'aqx-tam paper”, the number being “tam” and the numerical classifier being “'aqx”, i.e., “flat”

“three people” as “puma-t'útu lapanák”, the number being “t'-útu” and the numerical classifier being “puma”, i.e., “human”

“one snake” as “maqa-tam luu”, the number being “tam” and the numerical classifier being “maqa”, i.e., “flexible”

Note that even in the case of countable nouns the numerical classifier is added. Another example of classifiers that must not be confused with gender are noun classifiers. These are usually independent items (lexemes) that may or may not be attached to nouns themselves and signal specific semantic information about them. For instance, Aikhenvald (2019) gives the example of Yidiny, an Australian language in which the noun “boy” (“wurgun”) goes together with the “person” (“bama”) and “man” (“waguuja”) classifiers; so, to say “boy” they produce three different lexemes “bama waguuja wurgun” (literally: “person man boy”). Classifiers and gender systems can coexist in a language, as in the case of the Chinantec languages of Mexico, which have noun classifiers and a dual gender system based on animacy (see Aikhenvald, 2019; Foris, 1993; Palancar, 2014). However, these situations are rare (recent typological analyses find around 20 languages with more than one classification system; Fedden & Corbett, 2017; Sinnemäki, 2019). Numerous studies have shown that languages with gender systems tend not to have other classifier systems, and vice versa (e.g., Sinnemäki, 2019). It has been proposed that this is due to two factors: (1) economy and distinctiveness, by which a language in which a system of noun classification has been developed should avoid the development of another system that serves similar functions and may disrupt language processing, and (2) many studies point to classifiers as a possible first state of noun classification, from which gender systems are subsequently derived. In fact, classifiers are often thought to be the most common source of gender in languages (see Corbett, 1991, p. 136; Luraghi, 2011, pp. 450–452; Seifart, 2010, pp. 727–728), along with case and number agreement (Luraghi, 2011, p. 452).

On the other hand, if nouns can be classified in different categories that are not semantically instantiated, and if this classification is mainly syntactically inert (no agreement repercussions, Bermúdez-Otero, 2008), then these are inflectional classes or “declensions”, a morphological

feature by which nouns are classified according to arbitrary morpho-semantic or morpho-syntactic specifications (Kibort, 2008). Some examples are the nominative, accusative, dative, and genitive cases in German. Note that inflectional classes and gender usually cooperate, one determining the other (e.g., in Russian nouns from declension 1 are masculine, from 2 and 3 are feminine, from 4 are neuter; Aronoff, 1994; Doleschal, 2000). In sum, whereas other noun classifiers are semantically driven and inflectional classes are mostly grammatical, gender has a semantic basis but is g*rammaticalized*, the main difference residing in agreement.

Interestingly, the importance of agreement for the definition of gender is also shown in what is the basis for considering English a gendered language. More specifically, English lost its canonical gender system[3] between the 10th and the 14th centuries by attrition in the transition from Old to Middle English (for a comprehensive explanation on the decline of the English gender system, see Lass, 1992, Moore, 1921, and Faiß, 1989). Similar to German, English had a tripartite gender system (masculine, feminine, and neuter values). Apparently, attrition caused the loss of the agreement morphology reflected in adjectives, definite articles, and demonstratives, and subsequently the loss of gender in most of nouns (mainly inanimates). Ultimately, this triggered the restructuration of the assignment system into one based exclusively on sex (Audring, 2008; Moore, 1921). Grammatically speaking, gender now plays a minor role in English, since only some nouns referring to humans have semantic gender (woman, man; girl, boy; exceptionally also certain individuated nouns such as ships and cars, Siemund, 2013), whereas the rest of the nouns in the lexicon lack of any type of gender information. The existence of a semantic gender system based on sex provides the final piece of evidence for the canonical English gender system, which is now mostly anaphoric, in that it is only shown on agreement through pronouns (e.g., she/he/it; her/his/its). When pronouns are the only elements of agreement reflecting gender, these languages are said to have a "pronominal gender system" (Corbett, 1991, p. 170). Audring (2008) identified approximately 20 pronominal gendered languages in her typological study, among them Persian, plus languages from the Niger-Congo, Khoisan, Amazonian, Australian, and Austro-Asiatic families. The superiority of semantic gender on pronominal agreement can also be seen in what

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

[3] The term "canonical" gender system was proposed by Corbett (2006, 2012) to refer to the basic properties expected in a gender system (Audring, 2017). For instance, a gendered language would be expected: (1) to have each of its nouns belonging exactly to only one gender of the total available gender values, although there are languages with double-gender nouns (Corbett, 1991, pp. 181-182); (2) to have a semantic core; (3) to show gender agreement in the form of affixes and clitics and in more than one lexical category/syntactic domain.

is called “hybrid nouns”. Canonically, a gender system would assign each noun of a gendered language to a specific gender value. However, speakers treat some nouns, “hybrid” nouns, as belonging to two different gender values depending on the syntactic elements involved. These nouns are members of a certain gender value but can take the agreement markers of another value depending on the class of the word agreeing with them. In this sense, pronouns constitute the element most prone to change the agreement marker for semantic reasons (the likelihood of varying the agreement marker in function of the word class is well explained in the Agreement Hierarchy proposed by Corbett, 1979, 2006). Thus, a typical example of this phenomenon is that of the German word “Mädchen” (girl). In German, a common morphological gender regularity is that of the nouns ending in “-chen” being assigned the neuter value. Although “Mädchen” refers to a female human, semantic principles do not overrule here, and the noun abides strictly with the morphological tendency of assigning the neuter gender to that nominal ending. However, when it comes to the use of pronouns, “Mädchen” is commonly referred to with feminine personal pronoun “sie” (she) and not with neuter one, “es” (it).

The phenomenon of hybrid nouns is of particular interest to the current thesis in that it allows us to introduce two critical terms when talking about gender in the context of psycholinguistics: grammatical and natural gender. The case of “Mädchen” can be assessed from the perspective of a mismatch between those two kinds of gender types. “Mädchen” is classified as “neuter” following the same morpho-phonological assignment rules that inanimate nouns and some animate nouns do. Hence, “Mädchen” is a word with “grammatical” (arbitrary) gender. However, speakers sometimes seem to experience conflict, especially when using pronouns in which semantics appear to have a bigger role. Since “Mädchen” refers to a human being, as in the case of “Frau” (woman) or “Mann” (man), with pronouns speakers may opt for the semantically driven option, i.e., they treat “Mädchen” as having natural gender (and hence use “sie” [she], instead of “es”, [it]). In the following subchapter, we will explore these concepts in detail.

2.3 Grammatical and natural gender: Introducing the perspective of Psycholinguistics

USC UNIVERSIDADE DE SANTIAGO

From a psycholinguistic point of view, there is a distinction concerning the semantic roots of gender that must be taken into account: grammatical and natural gender. Grammatical gender is usually defined as a lexical-syntactic abstract property of nouns (Schriefers & Jescheniak,

1999; van Berkum, 1996), whose assignment is mainly arbitrary and does not depend on the intentions of the speaker. The only exception is indirectly linked to semantics, as it serves a discriminative function between words: the case of nouns that have the same form but different meanings, and the only differentiable parameter is their gender value. A clear example is found in Spanish: "cometa", when feminine, refers to a toy that is flown in the air on a long string (kite), but when masculine refers to a celestial body (comet)[4]. Besides, we may think here of those semantic fields correlating with one gender or another (e.g., colours are masculine in Spanish and Portuguese). Nevertheless, these are simply semantic regularities found in the assignment of grammatical gender to nouns within a certain language, but the gender value of these nouns has nothing to do with their meaning or the intention of the speaker. Thus, it is still abstract and arbitrary. Natural gender, however, is a lexical-syntactic and semantic feature that is directly dependent on the intentions of the speaker, leaving any trace of arbitrariness or abstraction behind. More specifically, natural gender is grounded in the semantic principles of gender assignment, but restricted largely to those languages in which the biological sex-split is the criteria for this assignment. It is, hence, not a mere semantic regularity found within gender (such as that of colours) but has its core in the real world. More specifically, the gender value of a noun with natural gender depends on the will of the speaker to refer to a male or female. This has significant consequences for the study of language processing, which is precisely the framework of this thesis.

Psycholinguistics studies the way language is acquired, represented, and processed in our mind. For a noun to be produced or understood, speakers have to recover it from their lexical memory (the "mental lexicon", the word store in our brain). In order to do this, they have to retrieve semantic, grammatical, syntactic, and morpho-phonological information, besides orthographic symbols and codes of articulation that allow language production to occur in its various forms (Garret, 1980). This process is called lexical access, i.e., the retrieval of all the information that represents a word in our memory. Lexical access involves the activation of multiple representational strata that contain different pieces of information. According to the most accepted structure of lexical access (Caramazza, 1997; Caramazza & Miozzo, 1997; Levelt et al., 1999; Roelofs, 1992), there are at least three of these strata. More specifically,

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

[4] As pointed out by Hickey (1999), whereas natural gender is guided by semantic needs, grammatical gender clearly is not: the semantic information given by the "masculine" and "feminine" gender is null in grammatical gender. These are labels of convenience which could be changed and still provide the basis for word distinction, as in the case of "cometa" (e.g., as in Swedish and Dutch, where gender values are the common and neuter values).

when producing a noun, a speaker must first encode the semantic features of the noun, contained in its "lexical concept" (Levelt, 2001). So, to produce the noun referring to a beaver in Portuguese, we must first encode the lexical concept "CASTOR", which will contain semantic features such as "animal", "rodent", "wild", "fur", and "river". Related lexical concepts will activate each other, thus, the activation of the lexical concept "CASTOR" will also activate other lexical concepts such as animal and otter (in Portuguese "ANIMAL" and "LONTRA"). For the production of the noun "castor" to occur, we must then encode a set of abstract and not-so-abstract properties that characterize the word that represents the concept of CASTOR, namely, its grammatical and syntactic features. Some authors argue that these properties are contained in an item called "lemma" (Levelt, 1989, 1992), a part of word's representation in the mental lexicon, the term first used by Kempen & Huijbers (1983). Hence, each lexical concept that is activated in our mind spreads activation to their corresponding lemma in the mental lexicon. Following this rationale, the concept of "CASTOR" will activate the lemma "castor", which specifies properties such as the word class (countable noun), number (singular), and grammatical gender (masculine or feminine in Portuguese). Finally, the lemma "castor" would activate its unique phonological code /castor/ as a set of ordered phonological (and morphological) segments (often known collectively as the lexeme). Having retrieved the corresponding phonological and/or morphological representation of that noun, these properties of the form are used to compute the corresponding articulatory gestures through a set of syllables that will be later converted into overt speech. The way that these parts of a word's representation and its associated features interact to produce or comprehend a word, as well as the time course of their activation and/or selection, will differ depending on the modeling of lexical access, as we will see in Chapter 5. Two points regarding gender processing must be borne in mind here:

1. It is widely accepted that gender is stored or computed in the lemma or syntactic-grammatical strata of lexical access (Caramazza, 1997; Caramazza & Miozzo, 1997; Levelt et al., 1999), as we have already noted. In the previous example, "castor" is an animate noun with abstract grammatical gender (beavers have a specific gender value in Portuguese, masculine); then, during the processing of the lemma, the appropriate gender value, masculine, would also be activated, along with its noun class and number.
2. Natural and grammatical gender will involve different strata for gender retrieval and will have different roles for the encoding of a noun.

USC UNIVERSIDADE DE SANTIAGO

The second point here underlines the major representational implications of whether or not gender is related to the meaning of a word. Nouns with natural gender have to specify their gender value through a semantic feature contained in their lexical concept that specifies the biological sex of the referent. For instance, for the activation of the Portuguese noun "menina" (female child, kid) to occur, semantic features such as "human", "kid", and "small" will be activated, such as in the case of "menino" (male child, kid), but the feature "female" will also be activated. This semantic feature referring to the biological sex of "kid" (represented in the stem "menin-") will allow the selection of the intended gender value, whose activation will be critical for the selection of the right gender morpheme, in this case "-a". Indeed, an important distinction between grammatical and natural gender is gender morphology. One of the differences that is usually highlighted between gender in nouns and gender in agreement (mainly adjectives) is the fact that the nature of gender is lexical for the former and inflective for the latter (Matthews, 1972). But note the following example from Spanish:

(3) "el armario alt-o/la estantería alt-a" (the tall closet/shelving)

"el niñ-o alt-o/la niñ-a alt-a" (the tall male/female kid)

In Example 3, it becomes clear that gender morphemes are commonly applied to nouns with natural gender (Shadid, 1997). There are exceptions, however, such as in the case of nouns with no gender value a priori and whose gender specification comes from the agreeing determiners: "el dentista/la dentista" ([the male/female dentist]; "dentista" has no gender value of its own). We could thus say that natural gender can be both lexical and inflective in its nature, whereas grammatical gender is purely lexical.[5] There are other examples in Spanish of gender morphemes applied to nouns, such as "-iz", "-esa", "-isa", or "-ina" (e.g., "actor" [actor] and "actr**iz**" [actress]; "tigre" [tiger] and "tigr**esa**" [tigress]). In German "-in" is added to form the feminine form of referents with natural gender ("der Student"/"die Student**in**" [the male/female student]). Grammatical gender is not marked through gender morphemes, but as we will see in the next chapter, it can be statistically linked to the phonology and derivational morphology of a language.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

[5] Some linguists argue that the type of inflection occurring with "menino/a" is the same as that with, for instance, the Spanish noun "banco/a" (same stem but different meanings). We will not focus on this debate, since it is not relevant for the object of study of this thesis; in any case, we do not think that these cases are cognitively similar.

As we have seen, in light of the previous Spanish examples, the gender value of a noun such as “armario” or “estantería” is irreplaceable, but the gender of “niñ-” (kid) is not determined a priori. This is another distinction to be made between grammatical and natural gender: the optionality of gender values. Following the definition of formal features made by Chomsky (1995), gender can be either intrinsic or optional. Nouns with grammatical gender have an intrinsic gender value, and hence “armario” cannot be feminine. Nouns with natural gender have an optional gender value, and so they can be either masculine or feminine.

A close inspection of the examples we have been using in this section reveals that the distinction between grammatical and natural gender is also closely linked to animacy. We have already said that every gender system seems to have a semantic core, in which the opposition between animate and inanimate beings is more or less emphasized. Still, even in those gender systems in which gender values are not assigned depending on the animacy of the referent (e.g., Romance languages with masculine and feminine gender values, such as French, Italian, Spanish, and Portuguese), the distinction between grammatical and natural gender depends on the degree of animacy. Indeed, the more we move to the left on the so-called Animacy Hierarchy (see Figure 2; Dahl, 2000), the more probable it is that we will encounter nouns with natural gender, that is, nouns in which the sex-split is relevant for grammar.

Figure 2

The Animacy Hierarchy



Note. This hierarchy, which has been described as a “hierarchic continuum” (Dahl, 2000), represents a bare minimum. Animacy has been shown to be at the foundation of many dichotomies in grammar. Grammatical phenomena follow this continuum by showing different generalizations or characteristics only above a certain cutoff point of the hierarchy. For instance, direct objects are more likely to be marked if their “degree” of animacy is higher (Dahl, 2000). Sometimes, to better locate this cutoff point, it is necessary to divide the continuum into more steps, for instance: animals with higher cognition and animals with lower cognition, or mammals and all remaining animals.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

In any case, there is not an exact cut-off point in gender systems with the distinction of grammatical and natural gender. While natural gender is restricted to nouns referring to animate entities (animate nouns), grammatical gender covers the whole spectrum of animacy. So, there

are nouns such as "niño/a", "jugador/a", "gato/a" (male/female kid/player; male/female cat) which have natural gender; and then there are nouns with grammatical gender such as "armario" (M), "estantería" (F), which are inanimate, "castor" (M) and "babuíno" (M) [baboon] which are animate but referring to animals, and "hombre" (M) (man) and "mujer (F) (woman), which are at the peak of the hierarchy (left extreme), since they are animate and depict human beings. All these nouns have grammatical gender, despite their degree of animacy, and hence feature an intrinsic (invariable) gender value which is not dependent on gender flections, and is of purely a lexical nature.

In sum, there are clear differences between grammatical and natural gender that must be taken into consideration, especially when studying the processing of language in the human mind. These differences are summarized in Table 1:

Table 1

Summary of the differences between grammatical and natural gender

	Grammatical gender	**Natural gender**
Relation to semantics	Null (arbitrary feature)	Full (semantically driven)
Relation with the intentions of the speaker	Null	Absolute
Choice of gender values	Intrinsic	Optional
Relation with word form (when existent)	Phonological	Morphological
Nature	Lexical	Lexical and inflective
Animacy	Inanimate + animate nouns	Animate nouns
Strata of lexical access	Lemma	Conceptual level + lemma

In the current thesis, we will focus on the representation and processing of grammatical gender during language production. Grammatical gender is the type of gender responsible for the "arbitrariness" appointed to this feature, along with the "puzzling" character referred to by many authors (Corbett, 1991). We will now focus on the relation between gender and word form, specifically between grammatical gender and phonology, to understand how this may shape gender acquisition and processing.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

3 GENDER TRANSPARENCY

As we have discussed above, one of the characteristics by which gender systems can be described is their degree of overtness (Corbett, 1991). Over the last three decades, psycholinguists have embraced the long-standing tradition in linguistics of looking at the link between gender and word form; however, rather than overtness, they tend to use the term "gender transparency". This term also refers to how evident gender values are in the form of the words, but critically, links overtness to ease of acquisition and to its potential repercussions for gender retrieval during the lexical access of nouns. From a psycholinguistic view, it does not matter how many form-regularities a gender system has if they do not influence cognition. In this thesis, we chose to describe phonological gender transparency as the core of the wider concept of gender transparency. After defining the phonological basis of a transparent gender system, we consider how the coincidence between phonological cues and gender morphemes may affect the whole degree of gender transparency of a language. All these elements are assessed taking into consideration the existing evidence regarding gender acquisition and processing. Finally, we propose a systematical assessment of the degree of gender transparency of languages through the adaptation of a series of standardized factors originally created by Audring (2019) in her analysis of the concept of relative complexity of gender systems.

3.1 PHONOLOGICAL GENDER TRANSPARENCY

The term of phonological gender transparency was popularized in 1995 through the work of Bates and colleagues after they observed that whereas most Italian nouns with natural gender exhibited gender morphemes, nouns with grammatical gender were partially related to phonology. They defined three types of nouns that could be found in Italian when looking at their nominal endings[6]. Nowadays, researchers have identified other languages in which these

USC
DE SANTIAGO
DE COMPOSTELA

[6] Referred by Harris (1980) firstly as "terminal elements".

types of nouns can also be found, namely Spanish and Portuguese. In all these languages, we can distinguish:

- Transparent nouns, which carry ending letters that statistically correlate with a certain gender value. For instance, in Spanish almost 100% of singular nouns ending in "-o" are masculine and 96% of singular nouns ending in "-a" are feminine (Caffarra et al., 2017; Harris, 1991). In Italian, according to a recent type-based count (Postiglione et al., 2014), the overwhelming majority of nouns showing the mapping '-a/-e' between the singular and the plural form are feminine (99.9%), while a similarly high proportion of nouns showing the mapping '-o/-i' between the singular and the plural form are masculine (99.97%).
- Opaque nouns, also called ambiguous, which end in letters that do not statistically correlate with any gender, such as the ending "-e" in Spanish and Italian (in Spanish, "torre" [feminine, F], tower; "puente" [masculine, M] bridge). For instance, in Spanish, approximately 30% of nouns are opaque (Harris, 1991). A similar percentage has been found in Italian (Postiglione et al., 2014); in this language, the most common opaque ending is "-e", and around 45% of nouns featuring this ending are masculine, with 43% being feminine.
- Irregular nouns are a reduced group of cases in which the transparent cues defined for a gender value are present in the opposite gender value, such as feminine nouns ending in "-o" in Spanish and Italian (e.g., "radio" and "mano" are feminine in both languages [radio and hand]). In Italian, these constitute only 1.5% of nouns (Postiglione et al., 2014); in Spanish, the number of irregular endings is higher for the feminine regular ending (4% of nouns ending in "-a" are masculine) than for the masculine cue (less than 1% of nouns ending in "-o" are feminine).

Transparent nouns are thus defined by positive statistical correlations between their gender value and their ending. These correlations arise for mere historical reasons (Villalva, 2008) and end up serving as cues for gender assignment. However, transparency defined as a simple correlation between nominal endings and gender values can lead us to think of other endings, usually derivational suffixes that also correlate with gender. For instance, in Spanish we can identify certain suffixes that normally serve another function but that indeed show a perfect correlation with gender. As seen above, the suffix "*-aje*" is added to verbs to form nouns that express action or certain rights that must be paid for, and those nouns are unequivocally

USC UNIVERSIDADE DE SANTIAGO

masculine ("aprender" → "aprendizaje" [learn → apprenticeship/learning]; "hospedar" → "hospedaje" [lodge → lodging/lodgment]). Other examples are the suffixes "-ción", which is used to form nouns from verbs and is fully feminine ("competir" → "competición" [compete → competition]), and "-ete", which is used to form diminutive, derogative or affective nouns from other nouns or adjectives and is fully masculine ("amigo" → "amiguete" [friend → buddy, in either a despective/confrontative or affective way, depending on the context]). We can identify at least another 13 suffixes in Spanish that correlate perfectly or almost perfectly with the feminine value, and at least 6 with the masculine value (Murillo, 1999; Shadid, 1997). However, these suffixes do not seem comparable with the endings "-o" and "-a" as defined by Bates et al. (1995, 1996), either in Spanish, or in Italian and Portuguese. In the following pages, we will aim to show how the "-o" and "-a" cues are unrivalled when compared to other types of endings, since they have a distinct role in language acquisition that seems to shape the way lexical access occurs in native speakers of Italian, Spanish, and Portuguese. In this way, we will lay the foundations for a more precise definition of what phonological gender transparency is, as well as for a systematical assessment of the general concept of gender transparency.

3.1.1 The distinct role of the transparent cues "-a" and "-o" in gender acquisition and processing

The literature on gender acquisition shows that Spanish children become very sensitive to morpho-phonological and syntactic gender cues from very early on. At the age of 24 months, Arias-Trejo and colleagues (2013) found that children acquiring Spanish were sensitive to the use of indefinite articles, and six months later to the use of definite determiners to disambiguate between masculine and feminine referents. However, they also found that this was true for nouns displaying the transparent cues for gender ("o" and "a"), whereas not for opaque nouns. Sensitivity to opaque nouns such as "pastel" (cake) and "flor" (flower) occurred only 6 months later. Hence, toddlers are faster to exploit the probabilistic co-occurrence between the gender of nouns and preceding articles or other elements of agreement when nouns exhibit the typical cues for gender. Similarly, it has been shown that children as young as 24 months are faster and more accurate in effectively classifying chains of letters (either words or non-words) as masculine or feminine when these exhibit the regular cues for gender (e.g., Lindsey & Gerken, 2012; Perona-Jara, 2015). On the same lines, infants as young as 30 months that are acquiring Spanish are able to refer to invented objects through novel pseudowords using the correctly

USC UNIVERSIDADE DE SANTIAGO

assigned gender following mainly the "-o" and "-a" endings (Arias-Trejo & Alva, 2013). The weight of the transparent formal cues in gender processing during childhood can be observed in later stages. Pérez-Pereira (1991) and Ogneva (2023) found that when the "-a" and the "-o" entered into conflict with other semantic cues of nouns with natural gender, children between the 4 and 11 years were better able to follow formal cues to categorize novel words according to gender. In fact, data on the youngest children in these studies suggest that phonological and syntactic cues of gender are realized prior to natural gender itself. This is not at all surprising if we bear in mind that up to three quarters of the nouns that Spanish children are exposed to during their first years of life comply with the formal o/a regularity according to the MacArthur Communicative Development Inventory (Jackson-Maldonado et al., 2003; Thal et al., 2000). In Italian, something similar happens, since a selective disadvantage for opaque nouns has been consistently found in healthy and language-disordered children and in deaf adults (Caselli et al., 1993; Leonard et al., 1992; Leonard et al., 1993).

An interesting and well-known phenomenon showing the critical role that the endings "-a" and "-o" play in gender acquisition is the numerous errors that children make as a result of the overgeneralization of these cues. For instance, Hernández-Pina (1984) reports overgeneralization of the feminine transparent cue, both in nouns and in other elements of agreement, in a 21 month-old toddler acquiring Spanish: "mota rota" (broken motorbike) instead of "moto (F, but irregular) rota" and "tierra azula" instead of "tierra azul" ("azul" [blue] is an invariable adjective across genders). Mariscal (2008), in a particular interesting case, found a child of approximately 22 months using "uno" (incorrect form of the masculine indefinite article: "a") in opposition to "una" (feminine indefinite article, "a"), in that he added the "-o" cue for the correct unmarked masculine indefinite article in Spanish "un" (e.g., "uno coche" instead of "un coche" [a car]). Fernández-Fuertes et al. (2016) followed gender acquisition in children acquiring Spanish as one of their first languages and found many examples of the overgeneralization of these cues, such as "una serpienta" instead of the correct form "una serpiente" (a snake), "un árbolo" instead of "un árbol" (a tree), and "uno azoto" instead of "un azote" (a spank). In Brazilian Portuguese, there are also reports of generalizations of the transparent cues in children around the age of 4 and 5 years (e.g., saying "dio" ["day"] instead of "dia" ["day", which is masculine but displays the typical feminine ending and thus is an irregular noun)], and overcorrecting irregular nouns, by which instead of saying "uma moto" [a motorbike], given that "moto" is feminine but ends in "-o" [irregular noun], a child

USC UNIVERSIDADE DE SANTIAGO

has been reported making errors of segmentation to transform “moto” in a transparent masculine noun: “um [M] amoto”; Figueira, 2001).

In sum, for those languages that display the simple gender cues “-o” and “-a” as defined by Bates and colleagues (1995), the literature on gender acquisition systematically finds that children are highly sensitive to these cues and to the syntactic distribution patterns of which they are a part. The degree of importance that these cues attain is high enough to induce generalization errors when children try to figure out the gender of opaque and irregular nouns and to establish agreement with the other elements of speech. A review by Hawkins and Franceschina (2004) concludes that French and Spanish children initially rely on phonological cues to realize concord, and only later does concord become a highly relevant factor for gender acquisition (for the inverse pattern, see Corrêa and Name, 2003, with Brazilian Portuguese). Indeed, the evidence discussed here contrasts with the situation with other perfectly correlated endings such as the derivational suffixes mentioned earlier (i.e., children have not been observed generalizing suffixes such as a “-aje” or “-ción” for the sake of gender assignment or agreement).

In terms of gender processing, after showing such a central role during language acquisition, it would be surprising that the phonological cues “-a” and “-o” would cease to play an active part in lexical access later in life. Indeed, a review of the literature shows that gender retrieval seems to be somehow shaped by phonological gender transparency in speakers of Italian and Spanish. Two types of studies can be distinguished: those assessing gender explicitly, and those assessing gender implicitly. Studies on the former rely on tasks in which participants have to press one or another key depending on whether the noun presented on the screen is masculine or feminine. These types of paradigms are called gender decision, categorization, or monitoring tasks (a term first used by Radeau et al., 1989). In terms of behavioural measures, researchers systematically find slower categorization times for opaque and irregular bare nouns than for transparent bare nouns (Bates et al., 1995, 1996; Hernández et al., 2004; Padovani et al., 2005). Studies using a functional magnetic resonance imaging (fMRI) to record the brain activity of adult speakers of Spanish and Italian during these tasks (Hernández et al., 2004, 2007; Padovani et al., 2005) found increased activity for opaque and irregular nouns in comparison to transparent nouns. The areas registering the peak of the activity were left BA 44 and BA 45. Left BA 44 has been described as reflecting memory retrieval processes and, when present, identification of morpho-phonological cues embedded in

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

the word itself; BA 45 is typically linked to syntactic operations and has been associated with the generation of determiners. The authors suggested that participants had been forced to:

(1) Make greater use of memory-based information to retrieve the gender of an opaque noun than that of a transparent noun due to the absence of phonological cues, which increased the activity of the left BA 44 in comparison to the case of transparent nouns.

(2) Produce a determiner when presented with opaque nouns as extra morpho-phonological help toward consciously making successful gender decisions, a strategy that had already been observed in another Italian gender-categorization study by Miceli and colleagues (2002) and which would have been responsible for the peak in the BA 45 area, absent for transparent nouns.

Importantly, evidence regarding the active role of phonological cues in lexical access has also been found in tasks involving the implicit assessment of gender. Facilitation in reading times for Italian bare nouns have been found when these are transparent in comparison to when they are irregular (see Experiments 1 and 2, De Martino et al., 2011; 2017). Participants had an advantage when reading feminine transparent nouns such as "casa" (house) compared to the reading of masculine irregular nouns such as "problema" (problem). These differences disappeared in the plural forms, in which the plural marker intertwines with gender and thus "-i" becomes a transparent plural gender cue for every masculine noun, regardless of its nominal ending in the singular ("pennarello" [transparent] → "penarelli" [marker pen/s]; "problema" [irregular] → "problemi" [problem/s]). In the plural, thus, both feminine and masculine nouns are transparent. The same advantage for transparent nouns was obtained when participants had to produce the singular (transparent) forms of the plural feminine nouns ("case" [houses] to "casa" [house]) in opposition to the singular (irregular) forms of the plural masculine nouns ("problemi" to "problema"; see Experiments 3 and 4). Studies exploring the processing of gender with fMRI measures in Italian speakers found the same facilitative effect for transparent endings when participants were asked to produce loudly the singular or plural forms of nouns. More specifically, Russo and colleagues (2021) observed increased activity in cortical areas associated with error monitoring and attentional control when the degree of ambiguity in the nominal ending was greater, especially for opaque endings, and even more so for irregular endings; interestingly, it was also the case for feminine endings when compared to masculine endings, this probably because in the singular the typical feminine cue "-a" has a prominent percentage of irregular nouns and in the plural it overlaps with the typically opaque inflectional

USC UNIVERSIDADE DE SANTIAGO

ending ("-e" as a feminine plural mark; and "-e" as a singular opaque mark, see Russo, et al., 2021). When, instead of bare nouns, participants were asked to repeat out loud noun phrases featuring articles and either transparent or irregular non-words, response times also showed a clear effect of facilitation for the former (Lindsey & Gerken, 2012).

Turning to lexical decision tasks in Italian (LDTs, in which participants are asked to decide if a chain of letters is a real word or not), De Martino et al. (see Experiments 5 and 6, 2011, 2017) found the same advantage for targets including feminine transparent and masculine irregular endings when analysing the number of errors for their singular forms, and the advantage was absent when plural (transparent) chains were presented to participants; note, however, that these facilitative effects for transparent nouns or non-words were absent for the analyses of the reaction times (RTs). In their paper of 2017, the authors hypothesized that the differences in their first two experiments and in their 2011 study may have been due to differences in gender (masculine vs. feminine) rather than differences in transparency, since they were comparing transparent feminine nouns and irregular masculine nouns. According to the Masculine Default Accounts (MDAs) and the available evidence (see De Martino et al., 2017), the masculine is the gender by default in Italian, and thus it can be automatically retrieved when a noun is encountered. In the case of irregular masculine nouns, after retrieving the masculine, the system enters into conflict with the gender cue present in the nominal ending, which pushes the inhibition of the masculine. This competitive process was seen to result in the inhibition of the feminine gender value, and in the selection of the masculine. The need for a double inhibition entails a disadvantage for the irregular masculine stimuli over the transparent feminine ones, for which the selection of the feminine is straightforward in a competitive account. To confirm the validity of this hypothesis, they replicated their reading aloud and LDT experiments with opaque targets, for which the masculine is automatically retrieved and accepted for masculine nouns but inhibited for feminine nouns. In this case, the feminine nouns are in disadvantage. There were no differences across the two types of targets, disconfirming the hypothesis regarding differences between masculine and feminine nouns, and supporting to the hypothesis regarding differences in the transparency of the nominal endings.

USC UNIVERSIDADE DE SANTIAGO

Interestingly, in another study with an LDT in Spanish, Urrutia and colleagues (2008) assessed the differential processing of transparent and opaque nouns (rather than of transparent and irregular nouns) and obtained higher RTs for the former. Transparent nouns were thus classified as real words slower than opaque nouns. Hence they argued that when explicit

decisions about gender are not required, transparent nouns involve higher levels of complexity than opaque nouns in terms of processing costs. Urrutia and colleagues stated that since opaque nouns are monomorphemic (e.g., "virtud" [virtue]) whereas transparent nouns are not, because they carry a thematic vowel (casa), then they consume fewer resources during lexical access. This interpretation, however, is questionable. Opaque nouns ending in "-e", which were also included in the list of opaque stimuli, are not monomorphemic: "-e" is also a thematic vowel. In fact, in Italian most nouns end in a thematic vowel (including opaque ones), thus they are not monomorphemic, and studies in which brain activity has been recorded through electroencephalographic (EEG) measures during reading tasks show increased activity for transparent nouns, which fits the findings of Urrutia and colleagues. Those studies were conducted mainly by Caffarra and colleagues with grammaticality judgement tasks and native speakers of Italian and Spanish (Caffarra & Barber, 2015; Caffarra et al., 2014, 2015). Their studies feature transparent, opaque, and irregular nouns that may or may not agree in gender with the preceding article. Native speakers of both Spanish and Italian showed transparency effects in early time windows (between 200 ms and 500 ms after the onset of the presentation of the noun) with a greater negativity associated with transparent nouns than with opaque nouns. They interpreted this as a reflection of the extra phonological encoding that transparent nouns go through. The transparency effect occurred before and without interaction with the agreement effect (i.e., unmatching gender article-noun pairs provoked a classical LAN-P600 biphasic pattern unaffected by the transparency of the noun). Overall, these results show that in Spanish and Italian the processing of gender-form regularities does not seem to be a post-lexical verification process but rather an active part of lexical access, at least during language comprehension.

In a recent study, also consisting of a grammaticality judgement task of noun phrases (definite article + noun) and fMRI measures, Quiñones et al. (2018) found that, in terms of RTs, gender mismatch conditions (in which the determiner and the noun had different genders) were harder for the participants when nouns were opaque than when these were transparent. The fMRI results showed that transparent and opaque nouns did indeed exhibit different patterns of neural activation. On the one hand, they found regions of increased cortical activity for transparent nouns, usually restricted to a left-lateralized network. On the other hand, they also found regions of increased cortical activity for opaque nouns, mainly involving a bi-lateral fronto-parieto-temporal network. Among the results, the clear increase in the activation of left

USC UNIVERSIDADE DE SANTIAGO

occipito-temporal regions during the processing of transparent nouns were shown to reflect morphological decomposition. Additionally, the supramarginal gyrus was also involved, and this was linked to a processing cost associated with decoding the redundant morphological information. As for opaque nouns, the high cortical activity of the posterior portion of the middle temporal gyrus and multiple bi-lateral regions appear to reflect a processing cost in the integration of the morphosyntatic information arising from the determiner.

In sum, the classical transparent cues for gender "-a" and "-o" seem to have a unique relation with gender that is not comparable to that of any other ending, even those that exhibit a perfect correlation with a specific gender value (e.g., derivational suffixes). These two cues have an overwhelming influence on the acquisition of grammatical gender, in that they statistically facilitate the identification of distributional patterns between nouns and other elements of agreement (mainly articles). The evidence not only suggests that they are part of the knowledge that speakers have of their gender system, but also that they continue to influence gender processing during adulthood, at least in language comprehension. Indeed, explicit tasks on grammatical gender show that participants rely heavily on these cues to respond, and opaque nouns seem to require extra analyses supported by determiners (which are sometimes produced silently) for a response to be given. Consequently, some authors suggest that the impact of transparent cues is only present when gender is required as an answer, but behavioural implicit tasks, as well as grammaticality judgement tasks with EEG measures, suggest the contrary. Furthermore, EEG studies show that the processing of transparent cues occurs early during lexical access and not in a post-lexical phase as a task-specific strategy. Hence, transparent cues seem to remain an active asset of gender retrieval. As for opaque nouns, when gender is not asked for explicitly in a task, it seems highly improbable that speakers process unnecessary articles to be able to access a noun and its gender. In this sense, the processing of articles is a better candidate for a specific strategy for certain tasks than the classical transparent cues.

3.1.2 Identifying cues for gender in other languages: The case of German and French

USC UNIVERSIDADE DE SANTIAGO

Having identified those factors that allow the phonological gender cues "-a" and "-o" to have such an impact in gender representation and processing in the abovementioned languages, we must also assess cases of gender regularities in other languages. In this way, we will be able to test the universality of these findings and understand to what extent there may be cross-

linguistic differences. Indeed, cues for gender have been identified in languages such as German and French, but these are not as simple as "-a" and "-o". If we take the "-a" and "-o" situation of phonological gender transparency as very close to the canonical situation of overtness defined by Corbett (1991), many languages would be situated at quite a distance from this. In this section, then, we aim to determine whether divergence from canonicity is relevant in defining the concept of gender transparency when it comes to ease of acquisition and gender processing repercussions. To do so, we must assess whether these form regularities in other languages play a comparable role as in Italian, Spanish and Portuguese.

3.1.2.1 The case of German

German is one of the most paradigmatic cases regarding the unpredictability of gender values (Hickey, 1999). There is an exhaustive literature on gender assignment with regard to German gender regularities, with the goal of demonstrating that claims as to the absolute arbitrariness of German gender assignment are wrong (Köpcke, 1982)[7]. In one way, the motivation behind these comprehensive studies can also be seen as evidence that straightforward cues like "-o" and "-a" do not exist in German and that the picture on gender predictability might be slightly more complicated than in Spanish, Italian or Portuguese. Indeed, Köpcke (1982) and Köpcke and Zubin (1983) found the astonishing number of 44 gender regularities in monomorphemic and extended monomorphemic nouns. This high number of regularities can be problematic. For instance, predictability increases as the lexical concreteness of the suffix attached to the syllable becomes greater. For instance, the endings "-en", "-el", and "-er" have been identified as being at the core of masculine gender overtness. However, when these extensions do not represent a concrete meaning, their capability of predicting gender diminishes. For example, "-en" as an ending of a substantival verb is almost always neuter "das Fahren" [the driving], whereas the ending "-er" has many exceptions when it is not part of an agent noun ("der Farmer" [the farmer] vs. "die Butter" [F] [the butter] vs. "das Zimmer" [N] [the room]). Importantly, the predictability of gender in German has also been shown to correlate with the sonority of the syllable rhyme, especially for vocalic suffixes. An example is the association of the suffix "-e" with the feminine value, which is probably the most relevant

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

[7] Bloomfield (1933) and Maratos (1979), along with other well-known authors such as Brugmann (1889) and Fodor (1959), were supporters of the Arbitrariness Hypothesis, by which gender would be a completely arbitrary feature of nouns assigned without any type of strategy or regularity. Consequently, it would have to be simply memorized for each single noun. German was often used as an example to reinforce this view.

nominal ending for gender prediction in German, especially when considering the whole range of cues in terms of similarity with the Spanish/Italian/Portuguese "-a" and "-o". In any case, even for the nominal ending "-e" there are many exceptions, namely in regard to loan words, where the stressing of the final vowel affects the degree of predictability (e.g., loan words in which the final vowel is unstressed may not be feminine ["das Entrée", N, "the entrance"]). Moreover, regularities on monosyllabic nouns also entail nominal beginnings in terms of phonology (words beginning with /kn/, /dr/, /tr/ and /s/ + consonant tend to be masculine). The intricacy of these rules goes even further, in that they combine phonemes, letters, and position within the noun itself: nouns ending on a nasal + consonant tend to be masculine, as well as nouns with two initial and two final consonants. All of these are examples of gender regularities in monomorphemic and extended monomorphemic nouns, but this leaves out a huge proportion of words that are polysyllabic. In polysyllabic words the ending "-e" again stands as feminine. There is, however, a considerable group of exceptions comprising nouns that start with "Ge-", a prefix that correlates strongly with the neuter value and leaves the regular feminine "-e" in an irregular situation (e.g., Gebärde [gesture], Gelache [laughter]). Nevertheless, the most regular endings for gender are mainly related to derivational suffixes, as seen for Spanish and Portuguese above. Hence "-heit", "-keit", "-ung", "-schaft" and "-ei", among others, are feminine, "-chen", "-lein", "-tum", and "ge-" are neuter, and "-ling" is masculine (Heidolph et al., 1984; MacWhinney et al., 1989). The picture gets even more complicated in that gender, number, and case intertwine, meaning that certain regular endings in certain combinations of these three features are also common for another gender in other combinations (e.g., the masculine cue "-en" may overlap the plural feminine termination: "die Frau", "die Frauen" [the women/women]).

Although in German, gender regularities form a body of complex correlations with many exceptions, studies on gender acquisition with German children show that they can rely on these patterns, especially when they are explicitly asked to assign gender to monosyllabic and polysyllabic non-words and real words (Mills, 1986; Szagun et al., 2007). In such studies, children score significantly above the level of chance when assigning gender following the degree of predictability of the word-pattern. Similarly, studies using explicit tasks to evaluate the knowledge that German adult speakers have of these regularities show the similar results. More specifically, gender categorization tasks of monomorphemic non-words through the use of definite articles and gender labels show faster responses and higher precision when these

USC UNIVERSIDADE DE SANTIAGO

cues are present (Köpcke & Zubin, 1983; Schwichtenberg & Schiller, 2004; Wegener, 1995). Evidence from an fMRI study conducted by Heim and colleagues (2005) suggests that, indeed, during gender categorization tasks, German speakers recover gender using a memory-based strategy, but some speakers who said they had used determiners to decide the gender of the word also showed an increased activity in the BA 45 area. The rise in the degree of activation of this area seems to be linked to the silent production of the determiner, as in the case of opaque nouns in gender categorization tasks with Italian and Spanish (see above). Unfortunately, in this study, the authors included all kind of nouns in the stimuli list, featuring multiple predictive values for gender (ranging from nouns with a high to a null degree of predictability) and the effects were not analysed according to this variability. In any case, participants that reported using the determiner strategy did not show differences in the strategy across words.

When the gender identification task is slightly less explicit, results change. For instance, if participants are asked to produce the correct definite article for each noun instead of selecting the correct gender or even the key with the correct article embedded, there are no differences between polysyllabic nouns of different degree of gender predictability in their derivational suffixes (Hohlfeld, 2006). When considering monomorphemic words in these less explicit tasks, the results are conflictive: Schiller et al. (2003) found behavioural differences in the categorization of regular and non-regular nouns to the benefit of the former, but the effect was absent in electrophysiological measures. A possible explanation advanced by the authors was that gender cues in German are not part of lexical access, but rather of a post-lexical access phase in which participants self-monitor to verify that the article they have chosen is correct, contrary to what has been shown for Spanish and Italian. This later effect would be reflected in RTs but not in EEG measures, which would be more sensitive to earlier effects during lexical access. The scant literature available, then, shows that these regularities are indeed part of the linguistic competence of German speakers. However, as noted by Hohlfeld (2006), due to the high quantity of merely probabilistic rules that must be taken into account to predict gender in German, speakers might learn not to rely on cues during online processing, this in order to avoid mistakes. Of course, German speakers have knowledge of these rules, since their capacity during childhood to identify them through statistical learning allows them to do so and helps them in the acquisition of gender as a feature. Consequently, as we have seen, they may use these regularities as a task-specific strategy when information about the gender of words is explicitly asked. However, the group of acquired regularities do not seem to constitute a very

USC UNIVERSIDADE DE SANTIAGO

efficient way of online gender processing during ordinary lexical access, and hence speakers might use them only when compelled to draw on this knowledge.

3.1.2.2 The case of French

French has also endured some quasi-accusations of arbitrariness in its gender assignment system (see Bosquart, 1998; Jacob & Laurin, 1994; Taft & Meunier, 1998):

> There seems to be no practical criterion by which the gender of a noun in German, French, or Latin could be determined: to define the meaning of the episememe 'masculine' in such a language would be simply to list the markers of masculine nouns and the nouns that belong arbitrarily to the class. (Bloomfield, 1933, p. 280)

Contrary to these claims, it has been shown that French is far from arbitrary when it comes to the presence of gender regularities. French is an Indo-European language from the Romance branch, like Italian, Spanish and Portuguese, with a dual gender system consisting of masculine and feminine. Although there are no cues as simple and frequent as the "-o" vs. "-a" distinction, correlations between the phonological form of the nouns and their grammatical gender have long been identified (Bidot, 1925; Mel'cuk, 1958). Indeed, Tucker et al. (1977) showed that gender can be predicted from the noun's ending in 84.5% of cases (considering 31,619 noun entries; see also the corpus of Lyster, 2006). However, there are many regularities, in addition to the exceptions and peculiarities. Consequently, the idea that these supposed rules and regularities could be the basis of gender assignment and processing in the French mind has been extensively criticized, since the rules seem to have an ad hoc character, sometimes grounded in extremely weak probabilities (Becket, 2010; Carroll, 1989; Warden, 1997). In any case, phonemes, suffixes and larger word units do correlate with gender, either phonologically or morphologically. Around 81% of nouns ending in /r/ are masculine, as are 93% of nouns ending in /ʒ / or /m/, and around 95 to 99% of words whose final phoneme is /o/, /o/, /y/, /ã/, or /k/. The endings correlating with the feminine gender are present in around 60 to 70% of the nouns that feature them (thus, 30 to 40% of nouns with these endings are masculine), except for /z/, for which 90% of nouns are feminine. The literature on regular endings that might mirror the role of "-a" and "-o" in other Romance languages reveals another troublesome aspect of French. For the feminine value, 70% of feminine words end in "-e", and "-e" is the usual marker for the feminine gender in natural gender and in agreement morphology (e.g., indefinite articles "un"

[M] vs. “une” [f], but see exceptions such as “père” [father, M]). However, when looking at the nominal ending instead of the gender value, 55% of nouns ending in “-e” are feminine, the remaining 45% being masculine. In other words: feminine nouns are frequently marked by the ending “-e”, but the ending “-e” itself does not correlate clearly with any gender. In fact, this is seen particularly through the creation of two terms within French research: ending-to-gender predictivity and gender-to-ending predictivity (Colè et al., 2003), turning the supposed “transparency” of French into something quite complex. Two situations enable such a peculiarity: (a) 59% of nouns in French are masculine, which creates a masculine bias and makes it easier for a highly feminine ending also to be highly masculine (Séguin, 1969); (b) it has been noted that regularities in French are better understood as larger units, primarily syllables, due to their main role in the perceptual segmentation of words in French (Dumay et al., 2002; Hardison, 1996). A computational simulation of gender attribution based on phonological cues developed by Matthews (2010) confirmed that, indeed, the predictability of French gender considering final syllables is significantly higher than the predictability of the final phoneme. In any case, when syllables come into play, the situation turns out to be even more intricate for those phonemes that correlate with a specific gender. One of many examples is the phoneme /õ/, which is feminine in 70% of cases. Whereas the ending /jõ/ is feminine in 91% of cases, /bõ/ and /rõ/ are masculine in approximately 87% of the time. Regarding the ending “-e”, again “-iene” is feminine, but “-age” is mostly masculine. Another complex and specific situation arises from the fact that French has a very opaque orthographic system, in contrast to Italian, Spanish, Portuguese, and even German. Consequently, there are many mismatches between phonology and orthography that can complicate the identification of the distributional patterns between nominal endings and gender when speakers acquire the ability to read and write. For instance, the ending /ã/, which is overwhelmingly masculine, refers to the final orthographical units of “-a”, “-and”, “-anf”, “-ang”, “-aon”, “-ens”, “-ends, “-empt” and “-ants”, among many others. Also, regarding the phoneme /e/, the orthographic sequences “-ée” and “-er” correspond to the phoneme /e/, nevertheless in the former it marks the feminine in 92% of cases and in the latter it always marks the masculine. Finally, another way of understanding form-gender regularities in French is through derivative suffixes, similar to those we described in Spanish, Portuguese and German: for French, -eur, -ie, ité, -icité, -esse, -itude, -étude, -ance, -ence and -ise, among others, are feminine suffixes, whereas -eau, -ot, -on, -eron, -in, -et, -illon, -icule and -isme are masculine suffixes.

USC UNIVERSIDADE DE SANTIAGO

Indeed, the distributional regularities between gender and the form of nouns in French create an intricate system full of many intertwined morpho-phono-orthographical predictive endings. It is not surprising, then, that findings in the literature of gender acquisition and processing are not clear regarding the role of "transparency" in French. The number of distributional relationships that have to be learned to use phonology to acquire and process gender similar to the way Italian, Spanish or Portuguese children do seems to be quite overwhelming (Matthews, 2010). The classic work of Karmiloff-Smith et al. (1979) initially conferred on French phonology a central role in gender acquisition, with syntax (particularly articles) determining gender attribution only after the age of four. Nowadays, this study, whose impact in the literature was once vast and continues to be widely cited (e.g., Matthews, 2010), has been roundly critiqued (see Boloh & Ibernon, 2013; Seigneuric et al., 2007). Probably the most striking fact here is that Karmiloff-Smith et al. (1997) could not replicate these previous results. The author adjusted her conclusions and stated that French children first learn sets of article/noun pairs as whole chunks of information and then start to extract the general system of morpho-phonological oppositions on word endings. The ability to generalize this system would be shown at four years of age, and articles would have a primary role on gender acquisition since the very beginning. Indeed, recent evidence has shown that even 25-month-old French toddlers are perfectly capable of using gender information on determiners when processing and comprehending subsequent nouns (Van Heugten & Shi, 2009). In any case, taken as a whole, the literature on French suggests that the role of phonology is unclear. Generally, studies exploring gender form regularities in French, in both children with a developing gender system and adults with a well-established one, have relied on gender categorization tasks. These studies show that native speakers of French are faster and more accurate when deciding the gender of visually presented nouns that show gender regularities in comparison to those whose ending is ambiguous (Desrochers & Brabant, 1995; Desrochers et al., 1989; Desrochers & Paivio, 1990; Holmes & de la Bâtie, 1999; Holmes & Segui, 2004, 2006; Taft & Meunier, 1998; Tucker et al., 1968, 1977). The results, however, are not as straightforward as they seem to be in Italian, Spanish and Portuguese. Seigneuric et al. (2007) showed that the ability to guess the gender of nonce nouns solely on the basis of endings emerges quite gradually in French, with scores starting at 60% at age 3 and culminating at 82% at age 9 (percentages referring to the number of correct nonce words classified according to the gender predicted by the nominal ending). Besides, gender decisions are not uniform for every

USC UNIVERSIDADE DE SANTIAGO

noun displaying endings that correlate with gender. More specifically, differences between nouns with predictive and unpredictive endings are consistently null for nouns that begin with a vowel (Arrivé et al., 1985; Desrochers & Brabant, 1995; Desrochers & Paivio, 1990; Desrochers et al., 1989; Taft & Meunier, 1998). The reason behind this seems to lie in the fact that when a noun starts with a vowel in French, the preceding definite article is invariable across genders (the neutral "l'" instead of the masculine "le" or the feminine "la"). Indeed, whereas children of the other Romance languages are extremely sensitive to the "-a" and "-o" cues of gender from the age of 24 months and regularly generalize them across nouns and even other elements of agreement, French children (and adults) seem to rely heavily on articles to make gender decisions. Participants from Taft and Meunier's (1998) study reported the use of definite articles for all targets, regardless of the predictivity of their nominal endings. Ayoun (2018) conducted a written gender categorization task with a large sample (n = 168) and found that participants displayed a better performance for consonant-initial than vowel-initial nouns. Notably, there was no correlation between participants' accuracy in attributing gender to the stimuli and the reliability of the phonological endings. Strategically, most participants reported that they used elements of agreement, especially articles, to decide between one gender or another. Moreover, when asked to categorize nouns according to gender by choosing between "un" or "une" (i.e., the masculine and feminine forms of the indefinite article, which do not vary in their form depending on the beginning of the next word), native French speakers always respond faster than when asked to choose between masculine and feminine (Desrochers & Brabant, 1995; Desrochers & Paivio, 1990; Desrochers et al., 1989; Muller-Gass et al., 2000). Muller-Gass et al. (2000) recorded the electrophysiological brain activity of a group of participants during a gender decision task involving nouns with highly predictive endings based on their final rhymes, and found strong evidence that French speakers rely mostly on articles to work out the gender of a noun. More specifically, the P3 (associated with stimulus evaluation) related to nouns categorized with indefinite articles returned to the base line, but the P3 related to nouns categorized with gender labels led to a second P3 of evaluation. Furthermore, the amplitude of the N2 component (related to the difficulty of a task in various paradigms) was significantly greater when labels were required compared to when the choice involved indefinite articles. This study seems to confirm that when faced with a classical gender decision task based on labels, native speakers of French tend to first build a phrase in agreement, and then verify the gender obtained with the available options for selection on the screen.

USC UNIVERSIDADE DE SANTIAGO

Interestingly, in Spanish the evidence suggest that this occurs only for opaque nouns, as described above. In conclusion, it seems that, similar to German, phonological cues might form a part of French speakers' competence. However, in French, these cues alone are insufficient to successfully carry out a gender decision task, and this, as seen, contrasts not only with Spanish, Portuguese and Italian, but also with German.

What is of more immediate interest to us here is what occurs with tasks that do not evaluate gender explicitly. As Boloh and Ibernon (2013) note, when the task does not involve metalinguistic operations related to gender, results seem to blur yet further the role of French nominal endings in gender processing. Desrochers et al. (1989), Desrochers and Paivio (1990) and Spalek et al. (2008) found little to no effects of facilitation related to transparency in their LDTs and word and picture naming tasks. Recent and very detailed studies with large samples conducted by Boloh and Ibernon (2010) support this. In an implicit gender attribution task with more than a hundred 4 to 10 year old children and adults, they found that facilitation through nominal ending regularities was not clear, especially for the feminine gender. Bare nonce nouns with suffixes of either high or neutral gender predictability were presented to participants in different lists. Participants then had to record a message indicating which of the previously presented objects should be allocated to each of three different bags. They performed masculine marked nouns at ceiling, with no differences in terms of the degree of gender predictability. Feminine nouns were clearly more challenging, but did not show differences between suffixes of different predictability: the number of erroneous assignments to gender was independent of the presence of any feminine cue. The authors suggest that the masculine gender would be set as the default value and the gender of feminine nouns would be learnt through co-occurrence with the feminine determiners. Hence, any trace of sensitivity to gender cues would be developed through intense exposure after the acquisition of grammatical gender had already taken place. This would contrast with Romance languages in which gender cues play a central role for gender acquisition. On these lines, Taft and Meunier (1998, third experiment) conducted a grammaticality judgement task in which adjectives preceding nouns with endings of different degrees of gender predictability could be correctly inflected or not across genders. Again, performance was significantly better for masculine than feminine nouns. In their 2012 study, Boloh and Ibernon replicated their previous results in two experiments presenting more than a hundred children of between 4 and 10 years with a list of determiner-nonce nouns pairs. This time the lists included indefinite articles preceding the nouns, and thus nouns with

USC UNIVERSIDADE DE SANTIAGO

feminine endings were preceded by masculine indefinite articles and vice-versa. They also used the same suffixes as in Karmiloff-Smith's (1979) study (e.g., "-ine", "-ette" or "-ise" for feminine, "-a", "-an" or "-ain" for masculine). Regardless of the age group, in assigning the gender of the noun participants consistently relied on the preceding article in the list that had been previously presented, especially in the case of feminine nouns, for which they almost never made suffix-based choices. In their 2013 study of natural gender, Boloh and Ibernon found that children were quite sensitive to masculine morphological cues referring to males, but not to feminine morphological cues for females (e.g., "-iere"). However, sensitivity to masculine natural gender morphemes only appeared at the age of 4, which is later than in Spanish and Portuguese. In conclusion, French children do not seem to exhibit high sensitivity to morpho-phonological cues.

Indeed, the masculine could be considered as default in French since it is attributed to several categories such as nominalized verbs, nominalized adjectives and participles, borrowed words (the majority of which are masculine, including English borrowings), and foreign nouns code-switched to French (which again are overwhelmingly masculine) (Ayoun, 2018). However, something quite interesting might also underlie these findings: note that in all their experiments, Boloh and Ibernon used feminine nouns that displayed either the final letter "-e" or the final phoneme "/e/" as nominal endings. Although "-e" would be the best candidate to mirror the role of "-a" in the feminine gender of the other Romance languages, it is actually quite problematic in that it creates an extraordinary situation of null ending-to-gender predictability in French. In their first experiment, Colè et al. (2003) conducted an LDT with bare nouns that had different degrees of gender predictability depending on the final bigram (e.g., "-et", "-ait", "-le" for the masculine; "-ise", "-ie" or "-te" for the feminine). They obtained an effect of gender markers that approached significance, but only for low frequency words, which were classified faster when endings had high predictive values than when they had low values. Importantly, the effect was restricted to masculine nouns, indicating that feminine nouns were not influenced by the presence or absence of gender regularities. Colè and colleagues were the first to point out that, rather than the larger bigram unit (e.g., "-et", "-te", or "-ie"), participants were perhaps influenced by both the phoneme /e/ and the letter "-e", which was the ending of all their feminine nouns. As we have seen, the apparently "feminine" "-e" is quite problematic in that it can fit opaque and irregular definitions. In their second experiment, in which all masculine and feminine nouns ended in this letter, they found no differences in

USC UNIVERSIDADE DE SANTIAGO

response times depending on the degree of predictability of the larger bigram units, and that gender did not interact with it. However, in their third experiment, using the same materials as in Experiment 2 but with noun phrases (definite article + noun) rather than bare nouns, and the eye-tracking technique, they found that the probability of inspecting the word ending was higher for stimuli with high predictive bigram-based endings than those with low ones. The effect, however, was far clearer for masculine nouns (.72 vs. .55) than for feminine ones (.58 vs. .53).

In sum, gender regularities in French, as in German, may be part of speakers' competence, but seem to stand out as quite troubling. Consequently, it is not surprising that when it comes to exploring the impact of these "cues" in lexical access (i.e., implicit tasks), it looks as if it is small or even null. All in all, we have identified three main factors contributing to this situation:

- There is a clear dominance of the masculine value over the feminine value in terms of statistics and cue reliability, which automatically creates an imbalance between values and disturbs the oppositional role of certain cues that would favour the acquisition of a dual gender system, as in "-a" (feminine) vs. "-o" (masculine).
- The most relevant cue for the feminine value, "-e", is also an extremely problematic one, since it is opaque or irregular in terms of ending-to-gender predictability.
- There is quite a mismatch between phonology and orthography, but orthography generally overrides purely phonological representations in terms of predictive strength (Lyster, 2006), which becomes a problem when we bear in mind that what makes regular cues of gender useful in terms of gender acquisition and processing is their impact from the very beginning of language acquisition. Orthography, which further blurs the role of many phonemes, is only acquired later in life, and adds new information on gender regularity that may distort the already established system and increase its complexity. Indeed, final phonemes and orthographic endings have been shown to interact in gender attribution (Lyster, 2006).

3.1.3 Conclusions on phonological gender transparency

Phonological regularities play a strong role in gender attribution in languages such as Spanish, Italian and Portuguese during childhood, and seem to play an active role in gender selection during life. Complex systems of gender regularities, such as those seen for German and French, are actually taken advantage of during childhood, and remain part of the speaker's competence, but the experimental and theoretical literature seem to suggest that they cease to

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

have any function during lexical access later in life. It might be surprising to the reader that children can identify, acquire and use all these complex rules unconsciously, with little to no effort. However, this is in line with current research on statistical learning. Indeed, studies on speech perception have shown that during the second half of their first year, infants become highly sensitive to the phonotactic patterns of words in their native language (e.g., Jusczyk et al., 1994; Szagun et al., 2007). Thus, children's special ability for statistical learning, along with the high frequency of occurrence of the phenomenon to be learned (i.e., distributional patterns of gender regularities and elements of agreement), may work conjointly toward a fast acquisition of gender. In sum, it seems that the specific learning mechanisms and types of information attended to at a particular time in development make the assignment of gender to nouns and the acquisition of agreement for young children an easy task (Szagun et al., 2007, pp. 469-470). Indeed, children can cope with this intricate system of regularities and internalize it because they attend to information that adults tend to ignore. This is supported by a vast amount of evidence on second language acquisition: when acquired later in life, gender systems stand out as one of the biggest problems for late learners (e.g., Hopp, 2013, 2016; Sabourin et al., 2006). Consequently, we should not expect adult speakers of French or German to rely systematically on an intricate system of regularities such as those just defined, especially if their capability to do so has reduced with age.

3.2 MATCHING OF GENDER PHONOLOGY AND MORPHOLOGY

We have established phonology as the base of gender transparency through the concept of phonological gender transparency. However, from a psycholinguistic point of view it seems that the overtness of the gender system cannot be understood as these mere correlations between values and final phonemes. The simplicity of such regularities in terms of number of letters and mono-functionality, the mapping of these regularities onto gender (i.e., one regularity per gender value seems to adjust better to the canon), the proportion of nouns with these regularities, and the coincidence between orthography and phonology, are some of the factors that seem to allow phonological gender cues to occupy a central role in gender retrieval and processing. However, one final factor that we have not yet mentioned, and which may be of significance in the unification of these regularities as representative of gender, is the matching between phonology and morphology. Having established phonology as the basis of gender transparency, we propose that a full definition of this concept must address the question of

USC UNIVERSIDADE DE SANTIAGO

whether there is a coincidence between the transparent phonological cues for gender with the gender morphemes for natural gender and for many of the elements of agreement in a language. For instance, in Spanish, the morphological gender system establishes "-a" as the immovable morphological marker for the feminine gender, and whereas "-o" is indeed the most representative marker for the masculine gender, it shares its role with "-e" and zero (no marker). Due to this situation, it is often said that the feminine morpheme is the only true marker for gender, in line with the marked vs. default gender hypotheses (see Prado, 1982). In Table 2, we have collected many examples of the Spanish natural gender system and agreement system in which we can confirm the ubiquitous nature of the gender cues "-o" and "-a", in a very similar manner to Italian and Portuguese:

Table 2

Morphological examples of the Spanish natural gender and agreement systems

	Masculine	**Feminine**	**English translation**
Natural gender	Sacerdot**e**	Sacedort**isa**	Prophet - Prophetess
	Duqu**e**	Duqu**esa**	Duke – Duchess
	Gall**o**	Gall**ina**	Roster – Chicken
	Juez **(-)**	Juez**a**	Judge (male) – Judge (female)
	Niñ**o**	Niñ**a**	Kid (male) – Kid (female)
	Chic**o**	Chic**a**	Boy – Girl
	Perr**o**	Perr**a**	Dog (male) – Dog (female)
	Gat**o**	Gat**a**	Cat (male) – Cat (female)
Agreement elements	El **(-)**	L**a**	The
	Un **(-)**	Un**a**	A
	Él **(-)**	Ell**a**	He/she
	Est**e**	Est**a**	This
	Es**e**	Es**a**	That
	Aquel **(-)**	Aquell**a**	That (further)
	Algún **(-)** /Algun**o**	Algun**a**	Some/any
	Ningún **(-)** / Ningun**o**	Ningun**a**	None
	Much**o**	Much**a**	A lot of
	Mí**o**	Mí**a**	Mine

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Tuy**o**	Tuy**a**	Yours
Suy**o**	Suy**a**	His/hers/theirs
Nuestr**o**	Nuestr**a**	Ours
Vuestr**o**	Vuestr**o**	Yours
L**o**	L**a**	it (pronoun)
Feliz	Feliz	Happy
Azul	Azul	Blue
Amarill**o**	Amarill**a**	Yellow
Bonit**o**	Bonit**a**	Pretty
Barat**o**	Barat**a**	Cheap
Content**o**	Content**a**	Happy, pleased
Rápid**o**	Rápid**a**	Fast
Roj**o**	Roj**a**	Red

Note. "*Feliz*" and "*azul*" are examples of gender-invariable adjectives; when adjectives agree in gender with the head noun they do so exclusively through the cues "-o" and "-a".

In their study on natural gender assignment in relation to word form, Boloh and Ibernon (2013) argue that, for the influence of semantic gender in grammatical gender or vice-versa to take place, children need to realize that the two subclasses of grammatical gender (masculine vs. feminine) mirror the ones in natural gender (male vs. female). Indeed, in languages such as Spanish and Portuguese, this process should be facilitated, because even if gender acquisition is initially guided through phonological cues rather than natural gender (see Pérez-Pereira, 1991), these cues coincide to a great extent with the morphemes for natural gender, as we have seen. This would lead us to think that such a coincidence may indeed constitute a link between both kinds of gender, whose entries may feed into each other, thus facilitating the acquisition of gender and further increasing the relevance of the endings "-o" and "-a" for the system.

The coincidence with many of the elements of agreement is such that the presence of the "-a" and "-o" morpho-phonological markers of gender can be extremely high, even in single sentences, as can be seen in these examples from Spanish, Portuguese and Italian (4):

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

(4) "L**a** niñ**a** pequeñ**a** de l**a** famili**a** recién llegad**a** es un**a** alumn**a** educad**a**"
"El hij**o** pequeñ**o** del vecin**o** del tercer**o** es un niñ**o** ruidos**o**"

"**A** menin**a** pequenin**a** d**a** famíli**a** recém chegad**a** é um**a** alun**a** educad**a**"
"**O** filh**o** pequen**o** d**o** vizinh**o** d**o** terceir**o** andar é um menin**o** barulhent**o**"

"L**a** bambin**a** dell**a** famiglI**a** appena arrivat**a** è un**a** studentess**a** istruit**a**"
"Il figli**o** piccol**o** del vicin**o** al terz**o** pian**o** è un ragazz**o** rumoros**o**"

[The little (female) kid from the newly arrived family is a polite student]
[The youngest son of the neighbor on the third floor is a noisy boy]

It is thus not surprising to find differences in how gender is acquired in languages like these, and also in others where the situation of gender regularities drifts away from the canonical "-a" and "-o". The ease of detection, identification, and active establishment of distributional patterns across elements of gender agreement must be higher than in the case of languages such as German:

(5) "Di**e** groß**e** Dam**e** der neu angekommenen Famili**e** ist ein**e** gebildet**e** Student**in**"
"Der jüngste Sohn des Nachbarn im dritten Stock ist ein laut**er** Junge"

[The great lady of the newly arrived family is an educated student]
[The youngest son of the neighbor on the third floor is a noisy boy]

3.3 STANDARDIZING GENDER TRANSPARENCY

We will now propose a systematical assessment of the factors that we think should define the degree of gender transparency of a language from the perspective of the potential repercussions for gender acquisition in first languages. In this way, we intend to standardize the definition of gender transparency and lay the foundations for the creation of an accurate transparency continuum on which any language can be located with a certain degree of precision. We will take as a basis here the notion of the relative complexity of a gender system as defined by Audring (2019). Considering that when the gender system is simpler the earlier its acquisition occurs (see Aduring, 2019, p. 39), four general factors should be analysed toward understanding how the characteristics of a gender system may influence the acquisition of gender. These factors are frequency, perspicuity, consistency, and monofunctionality. We will hence apply them to the analysis of transparency, taking into consideration the most ideal or canonical situation for a gender system: that in which the degree of gender transparency, and hence of simplicity, reaches its highest possible value.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

3.3.1 Factors to assess the degree of gender transparency of a language

1. **Frequency**: related to the number of times a child is exposed to a particular item or structure. The higher the frequency, the easier is acquisition.

Regarding phonology, we must consider the proportion of nouns displaying the cues for gender in the language, i.e., the proportion of transparent nouns in comparison to that of (1) opaque nouns and (2) irregular nouns. The higher the proportion of transparent nouns, the higher the number of times the child is exposed to these cues, and hence the ease of acquisition will be higher. Provided that there are non-transparent nouns, the higher the proportion of opaque nouns, the lower the proportion of irregular nouns, hence the higher the frequency by which transparent cues are indeed regular, and the higher the degree of predictability of these cues. Thus, a transparent gender system should only have transparent nouns; if there are also non-transparent nouns, it is better for the sake of transparency if those are opaque rather than irregular.

The number of times the child is exposed to those phonological transparent cues is also unavoidably tied to the number of number values and the number of phonological regularities per gender value. The higher the number of gender values, the less the exposure to each of the regularities associated to each of them. Also, the greater the number of regularities per gender value, the less the exposure to each one of them. Hence, the canonical situation defined by Corbett (1991) applies here: ideally, a gender transparent language would have a "one-to-one" mapping between form-regularity and gender value.

Importantly, since the basis of frequency is "opportunity for exposure", the variable "frequency of use" may also be relevant: how many nouns are transparent within the X most frequently used nouns in a language. Since we are considering first language acquisition, frequency of use should be measured from two perspectives: (1) frequency of use and oral exposure of nouns for toddlers (as in the MacArthur Communicative Development Inventory; Jackson-Maldonado et al., 2003), and (2) frequency of use in general, mainly through materials based on written sources (e.g., SUBTLEX; Keuleers et al., 2010; Soares et al., 2015).

In a broader sense, when considering morphology, we should ideally have the highest possible of elements of agreement and nouns with natural gender displaying gender morphemes. If these gender morphemes coincide with the phonological gender cues, the frequency of use and exposure to the forms "-a" and "-o" will also rise. The higher the number

USC UNIVERSIDADE DE SANTIAGO

of cases in which "-a" and "-o" forms are related to gender values, the more the times the children are exposed to the links between gender values and these cues, and the easier will be the statistical learning of the gender-distributional patterns within the language. In line with this, an interesting aspect noted by Audring is that semantically motivated gender might be acquired later in a language, since there is lower frequency in the input. However, this delay might be reduced due to the coincidence between gender cues and gender morphemes in natural gender (as in Portuguese, Italian and Spanish).

2. **Perspicuity**: related to the clarity and simplicity of the cues for gender; such cues are expected to work best when they are perspicuous and clear.

In terms of phonology, cues such as "-a", "-o" and "-e" are expected to work better than larger units, such as those defined for German ("-er", "-en", "-el") or even derivational suffixes in any language ("-ung" "-keit" and "-schaft" for German, "-aje", "-or" and "-ción" for Spanish). The clarity and simplicity of those cues may be affected by the mismatch between orthography and phonology: if different orthographical forms represent a certain phonological ending, once writing is acquired an extra grain of complexity may be added to the system, blurring the perspicuity of these cues. For instance, whereas in Spanish the ending phonemes /a/ and /o/ are presented exclusively by the orthographic endings "-a" and "-o", in French the ending /ã/ refers to the final orthographical units of "-a", "-and", "-anf", "-ang", "-aon", "-ens", "-ends, "-empt" and "-ants", among others.

Additionally, with regard to the matching between phonology and orthography, an important concept here is that of syncretism: if the same element is a cue for different functions, the transparency of this cue as a whole will be negatively affected, as also will its usefulness for language acquisition and processing. The mismatch between orthography and phonology is a candidate for creating situations of syncretism, directly affecting perspicuity, and hence increasing difficulty in acquisition. For instance, in French the phoneme /e/ corresponds to the orthographic sequences "-ée" and "-er", yet the former marks the feminine and the latter the masculine. Hence, this is also related to transparency irregularity. As we have seen, irregular nouns affect frequency, because they diminish the proportion of transparent nouns in the language and the frequency by which the cues for gender represent the expected value. However, irregular nouns may also create a situation of syncretism that impedes the perspicuity of gender cues. The greater the number of irregular nouns, the lower the predictive value of the

USC UNIVERSIDADE DE SANTIAGO

cue as well as its degree of representativeness for a certain gender (e.g., the ending "-e" in French, which is syncretic for the feminine and the masculine).

Syncretism in the agreement system of a language may also affect the capacity of children to identify the distributional patterns between phonological gender cues and the elements that go with the head noun. For instance, in Spanish, the seemingly masculine definite article "el" is used as feminine when the next feminine noun begins with a stressed "a" (e.g., "alma" [soul], "arma" [gun], "águila" [eagle]), and so is syncretic for the masculine and the feminine, diminishing its capability to predict the masculine value. In this sense, the existence of cases in a language tends to increase the complexity of the gender system and decrease its transparency, since it usually promotes syncretism. An element that is regular in one case may also be regular in another, creating a situation of irregularity for the element itself. For instance, the German definite article "der" is syncretic for nominative masculine, but also for genitive feminine.

3. **Consistency**: related to the fact that the most transparent cues for gender are also the most consistent, and so an ideal cue has a unique form that consistently represents a particular gender value, and in an ideal transparent system a particular gender value is consistently represented by the same cue.

Paraphrasing Audring, the validity of one cue is directly related to its predictive power, which depends on the consistency to which it is associated with a certain gender value. The irregularity affecting one potentially predictive nominal ending decreases its consistency, and hence it decreases its transparency and the ease of acquisition of the cue and its corresponding gender value. In this sense, the French phonological nominal ending "-e" is a clear case of inconsistency: almost fifty percent of the nouns displaying this cue are masculine, despite the fact that most feminine nouns end in "-e". This is directly related to the syncretism we described above: in a situation of syncretism the cue is representative of two different genders, which overall increases the inconsistency of this gender cue.

Importantly, when various gender cues represent the same gender value, a situation of inconsistency arises by which the gender value cannot be systematically represented by a unique cue. In this sense, it is expected that the transparency and ease of acquisition of a "one-to-one" mapping situation (e.g., "-o" to masculine; Spanish, Italian, Portuguese) is higher than that of "various-to-one" mapping (e.g., non-palatal consonants to masculine; Russian). This matter can also be understood inversely, that is, in terms of the gender value in relation to a certain cue (gender-to-ending predictability). When a gender value is consistently represented

USC UNIVERSIDADE DE SANTIAGO

by one cue, its transparency is increased, as well as the ease of acquisition of this value. In this sense, the fewer the phonological regular endings associated with each gender value and the fewer opaque and irregular nouns, the higher will be the consistency of one gender to a specific gender cue, and the easier the acquisition of this gender and its corresponding cue. For example, Spanish contains a huge proportion of masculine opaque nouns when compared to feminine opaque nouns, and hence the consistency of the masculine gender value in relation to the cue "-o" is lower than that of the feminine gender value in relation to the cue "-a". More specifically, around 60% of masculine nouns end in "-o" and 77.5% of feminine nouns end in "-a" (Teschner & Russell, 1984).

The matching between phonology and orthography is also relevant here, since the greater the orthographical representations of a certain phonological cue, the less is the consistency of the orthographical form of the cue, and hence, in face of this inconsistency, the higher the difficulty for the speaker to acquire the pattern of associations between this cue and gender.

In terms of morphology, the ideal situation is when gender morphemes coincide with the phonological cues for gender, since this creates the most consistent link between morpho-phonology and gender values and hence increases its gender transparency and its ease of acquisition. Conversely, the presence of "deviant" gender morphemes diminishes the consistency by which a particular morpheme represents gender, which increases the opacity of the system and the difficulty of acquisition. As we have seen in Spanish, feminine agreement words are far more consistently represented, as they rely on the morpheme "-a", whereas the masculine relies on the morpheme "-o" but also on the cues "-e" and zero. It is worth mentioning here two fundamental concepts originally defined by Mills (1986), both of which are relevant for transparency as suggested by Caffarra et al. (2017): the concepts of availability and reliability, which involve the factors of frequency, perspicuity (syncretism), and consistency. Availability refers to the number of nouns displaying a certain form-function correspondence in a given language (i.e., transparent nouns minus non-transparent nouns). Reliability refers to the consistency of the relation between form and function in a given language (i.e., transparent nouns minus irregular nouns). The strength of a form-function mapping (i.e., cues for gender values) depends on both the availability and reliability of this mapping. Hence, when a probabilistic pattern covers a high number of items (high availability, large proportion of transparent nouns) and the number of exceptions is low (high reliability, low number of

USC UNIVERSIDADE DE SANTIAGO

irregular nouns), it provides a strong distributional cue, which is considered to be influential in language acquisition and comprehension (MacWhinney, 1987, 2004).

4. **Mono-functionality**: simple and thus transparent gender cues should display no other linguistic function than to statistically correlate to gender or to morphologically mark gender. Polyfunctionality, then, increases the degree of relative complexity of the potential cues for gender and decreases their usefulness for gender processing.

Significant situations of polyfunctionality are those in which gender intertwines with other features such as number, case, and inflectional class. According to Audring (2019), any kind of polyfunctionality affects both clarity and consistency. In Italian, for instance, gender and number intertwine in the plural, as "-i" is the masculine plural marker and "-e" is the feminine plural marker. However, case seems to be the feature that may exponentially increase the relative complexity of the form assignment system. In fact, Eichler (2013) observes that German gender is acquired later than French, Italian or Spanish gender, and attribute this to the influence of case.

Finally, polyfunctionality may also explain why derivational suffixes, which comply with the simple definition of "strong correlation between a nominal ending and a certain gender value", do not seem to have the same kind of role in gender acquisition and processing as the cues "-o" and "-a". Despite being cues for gender (although not simple ones), their function is something else: that of creating new words and assigning new functions to these words (e.g., to reflect action, add a derogatory nuance to a noun, etc.). This may also explain why they are not over-generalized in terms of gender.

3.3.2 Definition of gender transparency and summary of the determining factors

We will define the degree of gender transparency of a language as the extent to which the form of words plays a role in the acquisition and processing of the grammatical gender of nouns depending on the frequency (or availability) and consistency (or reliability) of the associations between certain word patterns with each gender value, as well as their perspicuity, syncretism, and mono-functionality. In Table 3, we present the summary of the abovementioned determining factors of gender transparency.

USC UNIVERSIDADE DE SANTIAGO

Table 3

Factors to operationalize the degree of gender transparency of languages

Factor	Modality	Subfactors	Ideal
Frequency	**Phonology**	Proportion of gender cues in nouns	Highest
		Number of gender cues per value	Lowest (one per gender value)
	Morphology	Proportion of gender morphemes in nouns and agreement elements	Highest
		Number of gender morphemes per value	Lowest (one per gender value)
	Both	Frequency of exposure and use during childhood	Highest possible number of nouns displaying those cues
		Overall frequency of exposure and use	Highest possible number of nouns displaying those cues
		Number of gender values	Lowest
Perspicuity	**Phonology**	Simplicity	Simplest
		Syncretism	Ending-to-gender predictability equals gender-to-ending predictability
		Matching between phonology and orthography	Absolute (orthographically transparent language)
	Morphology	Simplicity	Simplest
	Both	Syncretism (in agreement elements)	Agreement elements represent only one value
Consistency	**Phonology**	Proportion of irregular nouns	Lowest
		Proportion of opaque nouns	Lowest
		Mapping of gender values to gender cues	One-to-one
	Morphology	Coincidence between phonological gender cues and gender morphemes	Absolute
Monofunctionality	**Phonology**	Intertwining with other features	No intertwine
		Gender cues are monofunctional	Cues are not derivational suffixes

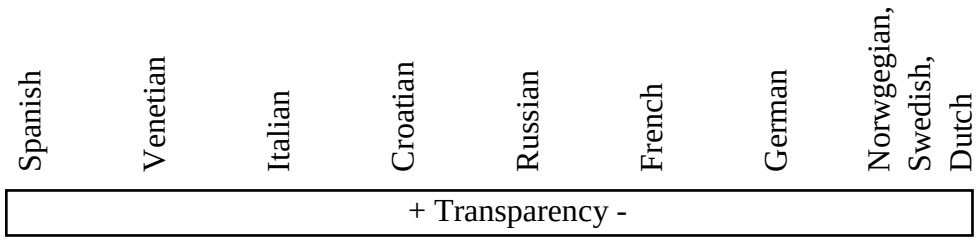
Note. The ideal situation reflects the highest degree of gender transparency, which in terms of relative complexity will provide the highest degree of simplicity and ease of acquiring a gender system. Note that one unique situation may affect more than one factor (e.g., the existence of irregular nouns affects the proportion and frequency of use of gender cues, their syncretism, and consistency in terms of gender-to-value and value-to-gender linking).

The use of these factors to operationalize the degree of gender transparency of languages should allow the creation of a transparency continuum in which languages can be precisely located. Indeed, Küpisch et al. (2018) were the first to propose some kind of continuum of transparency with Spanish, Venetian, Italian, Russian, French, German, Norwegian, Swedish, and Dutch (see Figure 3). This continuum was later updated by Velnić (2020), who included Croatian. Yet although an interesting proposal, it lacks the standardization that is desirable to be able to conduct proper predictions on language acquisition and processing depending on the

location of the target language/s on that continuum. We believe that the factors described above will allow so to happen, improving the possibilities and quality of future research.

Figure 3

Proposal of a continuum of gender transparency



Note. Adapted from Kupisch et al. (2018) and Velnić (2020).

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

4 MODELS OF REPRESENTATION OF GRAMMATICAL GENDER

The representation and processing of grammatical gender during language production has not been a primary focus of attention for theoretical and computational psycholinguistics. As a result, the presence of grammatical gender in the modelling of lexical access is rather low and limited, as evidenced by the fact that gender transparency has not been considered in any of the general models of language processing. In fact, some influential theoretical approaches in the literature have barely acknowledged the existence of gender (e.g., Dell, 1986, 1990). However, among the classical models of language production, two have included this feature within their tenets: the Word-form Encoding by Activation and VERification model (WEAVER++; Levelt et al., 1999) and the Independent Network model (IN; Caramazza, 1997). Both of these assess lexical access within the mind of native speakers and locate grammatical gender at a level distinct from semantics and word form, namely at the grammatical and syntactic stratum of word encoding. Within bilingualism, no model of language production has considered gender and its properties in its postulates. In this chapter, we will assess the structure and functioning of lexical access as proposed by the two abovementioned models, and will explain in detail how grammatical gender is conceived in both of these.

4.1 WORD-FORM ENCODING BY ACTIVATION AND VERIFICATION MODEL (WEAVER++)

The WEAVER++ model (Levelt et al., 1999; Roelofs, 1996a; 1996b; 1997c) was developed through data mainly on naming tasks measuring RTs. It features four levels of language processing: lexical concepts, lemma selection, lexeme encoding, and the phonetic encoding of words[8]. At the first level, every lexical concept is represented in a non-decompositional[9] way,

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

[8] The notion of lemma and lexeme distinction was first introduced by Kempen and colleagues (Kempen & Hoenkamp, 1987; Kempen & Huijbers, 1983).

[9] Non-decompositional or non-componential representations cannot be divided in lower-level units, and hence non-decompositional lexical concepts do not represent meaning through sets of semantic features but as whole chunks of information.

and the processing leading up to the activation of each one of these lexical concepts is called "conceptual preparation". Thus, if a speaker intends to refer to a beaver, he will activate the lexical concept "BEAVER". Importantly, every lexical concept is connected to semantically related concepts, so "BEAVER" is connected to other lexical concepts such as "OTTER","ANIMAL" and "RODENT".

Lexical selection starts with conceptual encoding and ends with the retrieval of the word lemma, which occurs at the lemma level. More specifically, an active lexical concept at the conceptual level activates its respective lemma node, and the most activated lemma is selected. The only exception is the selection of function words on purely syntactic grounds. Upon lemma selection, the proper syntactic environment for the word becomes available. This includes the encoding of grammatical features. For instance, the selection of the lemma "mala" (purse) in Portuguese will make available properties such as noun class (noun), number (singular), and grammatical gender (feminine).

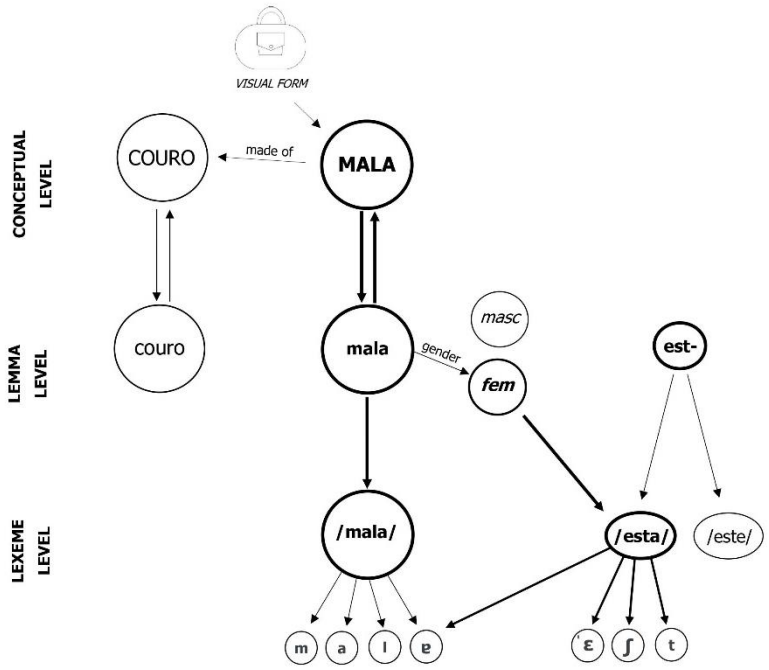
After the selection of the lemma has taken place, phonological/articulatory processing occurs at the lexeme level or form stratum. In this level, the speaker must retrieve the word's morphological makeup, its metrical shape, and its segmental makeup. If the lemma intended to be produced is "mala", the speaker will activate the lexeme /mala/. However, if the lemma were to be "malinha" (little purse), then the speaker would access two morphemes /mal/ and /inha/. The speaker would then retrieve the metrical information: "mal-" is a monosyllabic root; "-inha" is a disyllabic morpheme stressed on the first syllable, which cannot be an independent phonological word. Finally, the segmental spelling would occur for /mal/ and /inha/ (/m/, /ɐ/, /l/; /i/, /ɲ/, /ɐ/). The syllabification of the phonological form of the word occurs later due to its dependency on the phonological environment in which it is inserted.

Finally, Levelt and colleagues make a distinction between phonetic encoding and articulation, but they address it only very briefly, since it is not their aim to explain effects related to this level of language processing. The phonetic encoding involves the computation of the phonological word's gestures that serve as the basis for articulation. Each phoneme (such as /m/) would activate all syllables in which it is contained. Each of those syllables compose the gestural scores. A statistical procedure would now favour the selection of a certain gestural score among all active gestural scores, i.e., the one corresponding to the morpho-phonological form that has been encoded. After this, articulation takes place, and so the phonological word's gestural score is executed by the articulatory system.

USC UNIVERSIDADE DE SANTIAGO

Figure 4

Simplified structure of lexical access according to WEAVER++



Note. Processing of the Portuguese noun phrase "esta mala" (this purse). Bold nouns and thicker lines/circles represent selected nodes; non-bold and finer lines represent activation. Adapted from Levelt et al. (1999).

The theory is modeled in terms of an essentially feed-forward activation-spreading network represented in Figure 4 with the three strata of nodes that have consistently found support in the literature: the conceptual stratum, the lemma, and the lexeme or word-form stratum (Dell, 1986; Dell & Reich, 1981; Stemberger, 1985). Each stratum consists of nodes that accumulate activation. In the conceptual stratum, there is one concept node per lexical entry, as well as conceptual links (e.g., "MALA" [concept node] <<is>> [conceptual link] "RECEPTÁCULO" [concept node, -container-]; <<made of>> [conceptual link] [COURO] [concept node, -leather-]). A word's meaning is represented by the total of the lexical concept's labeled links to other concept nodes. Each concept node is linked to their respective lemma node in the lemma level. For instance, the concept node "MALA" is linked to the lemma node "mala" in the lemma node. The lemma node "mala" receives the full amount of "MALA" activation, however, since other concept nodes were also activated through conceptual links with the concept node "MALA", other lemmas receive a proportion of this activation (e.g., concept node "COURO" activates lemma node "couro"). After the system verifies that the link between the most activated lemma

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

node and the conceptual entry is correct (i.e., verification mechanism), the correct lemma is selected ("mala"). The selection of the lemma entails the activation of a series of syntactic and grammatical features. More specifically, each lemma node is linked to a group of relevant syntactic property nodes and to labeled links between them (e.g., "mala" <<number>> "singular"). The total of syntactic nodes connected by labelled links to a lemma node represents the word's syntax and grammar. After the selection of the lemma and the associated syntactic nodes, activation spreads to the word-form level or lexeme, which contains multiple morpheme and phoneme nodes. Only selected lemmas (and not active lemmas) become phonologically activated. Each morpheme and phoneme is linked to the relevant segment nodes ([m] [a] [l] [ɐ]). There is also a set of procedures that generate a phonological word's syllabification given the syntactic and phonological context and the syllabification rules of the language.

Some other important properties of the model remain:

- Lexical selection involves the conceptual and lemma strata, but lexical access consists of all steps: lexical selection and morpho-phonological encoding.
- There is feedback between the lemma level and the conceptual strata, but there is no feedback between the lemma and the lexeme levels, and hence active units in the form stratum do not affect the normal processing at the lemma level (recent updates do accept some kind of feedback activation through a monitoring system that had been conceived in the original proposals of the model but that had not been addressed in detail, as we will address later).
- Lexical access is a sequential process, i.e., it is strictly serial and there is no free cascading of activation through the lexical network (recent updates propose the existence of a limited cascading of activation in very specific experimental situations, as we will address later).
- There are no inhibitory links in the network, but nodes compete for selection within each stratum.
- Competition in lexical access occurs in two phases (Levelt et al., 1999; Roelofs, 1992, 2003; Roelofs & Piai, 2015): in the first phase, the activation of the noun to be produced must exceed the activation reached by that of all the other nodes by a critical amount, this called the critical difference; once the critical difference is achieved, the second phase consists of the actual selection of the target as a random event in time, whose probability of occurring is determined by the so-called Luce ratio (i.e., the activation of

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

the target divided by that of the sum of the activation of the competitors and the target itself, Luce, 1959; Roelofs, 1992; p. 118; Roelofs & Piai, 2015, p. 421).

We will now address the specific approach to grammatical gender selection as conceived by the WEAVER++. First, as we have seen, grammatical gender is located at the syntactic and grammatical level of encoding, otherwise referred to as lemma, in the form of nodes: one node per gender value. For instance, in Portuguese the language system is expected to entail two nodes: the masculine node and the feminine node. All noun lemmas point to one of the gender nodes. Yet these gender nodes point to multiple agreement targets, providing access to determiners, pronouns, and gender inflections. Hence, as represented in Figure 4, within the noun phrases "esta mala" and "este bolo" ("this purse" and "this cake"), the Portuguese noun "mala" points to the feminine node, which in turn points to the phonological form of the feminine demonstrative "esta" (this), and the Portuguese noun "bolo" points to the masculine node, which in turn points to the masculine demonstrative "este" (this)[10]. These gender nodes do not have backward pointers to the noun lemma, so the link between noun lemma nodes and gender nodes is unidirectional. Nodes are thought to function by accumulating activation. Importantly, the authors propose a distinction between activation and selection of the syntactic and grammatical nodes located at the lemma level. More specifically, once the lemma node is selected, other linked grammatical and syntactic nodes, such as gender nodes, receive part of the activation that had reached the lemma node. However, activated information is only selected when necessary (Roelofs, 1992, 1993). Thus, the specific gender value of a noun is only selected when a context of agreement is present, because in that case the gender of the noun must be computed for the form of the words which have agreement to be determined (e.g., definite and indefinite articles, demonstratives, adjectives, etc.; Levelt et al., 1999, p. 14; Roelofs et al., 1998, p. 224). So, for instance, when a speaker of Portuguese produces "mala" [purse] or "mala azul" [blue purse], in which "azul" is an invariable adjective across gender values, the feminine node would not be selected, only activated. Conversely, when the speaker produces "a mala" [the purse] or "mala amarela" [yellow purse], in which both elements of agreement vary across gender values ("a" [F] vs. "o" [M]; "amarela" [F] vs. "amarelo" [M]), the feminine gender node is selected.

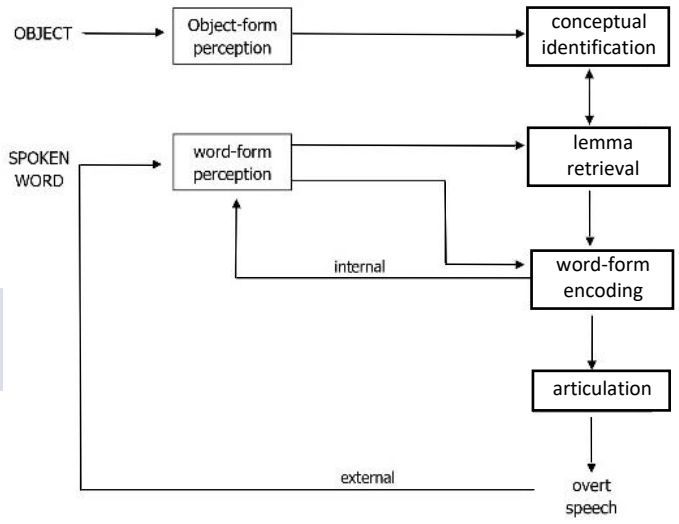
USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

[10] Some authors propose that gender nodes provide access to the lemmas of the agreeing targets rather than to their phonological forms (e.g., La Heij et al., 1998).

In subsequent updates of the WEAVER++ model, some clarifications have been made regarding the serial nature of the original proposal. On the one hand, WEAVER++ has explored in greater detail the comprehension-based self-monitoring system originally proposed in 1999. The theory conceives the idea of self-monitoring since the speaker can monitor their overt speech output and their internal speech through the representations that form during speech encoding (Levelt, 1983). More specifically, according to WEAVER++ the conceptual and lemma strata are shared between production and comprehension, but the input and output representations are separated for the lexeme level. Hence, within the production system, after lemma retrieval takes place, spoken word planning is strictly a feedforward process (Roelofs, 1997). However, the authors propose that there are two self-monitoring routes operating through the speech comprehension system, an internal route and an external route (see Figure 5). The authors maintain this idea in the outcomes of multiple computational simulations and in studies showing that the same neural structures are active during self-monitoring and speech comprehension (Hartsuiker & Kolk, 2001; McGuire et al., 1996). The external route involves the listening to self-produced overt speech, whereas the internal route includes monitoring the speech plan and feeding a rightward incrementally generated phonological word back into the speech comprehension system. So, the model accounts for the existence of a kind of connection between the sounds to lexical items (Roelofs, 2004; p. 564), but this is made through a comprehension system for self-monitoring purposes to explain task-based effects (mainly effects mixing both semantics and phonology, see Roelofs, 2004).

Figure 5

Representation of the two self-monitoring routes of lexical access in WEAVER++



USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Similarly, in the latest updates of the model, the authors have admitted the existence of cascading processing between the lemma and the lexeme, but only in a very weak form, and again limited to certain task situations. This cascaded activation, though, does not seem to depend on the intention of the speaker to produce the lemma (Roelofs, 2008). Evidence for these updates comes from various studies in which phonological facilitation and semantic interference effects were obtained in picture-picture interference tasks. In these tasks, participants are asked to name pictures while ignoring other distractor pictures that may or may not have any type of relation with the target pictures. Morsella and Miozzo (2002) and Navarrete and Costa (2005) observed that pictures were named faster when the distractor picture was phonologically related than when it was unrelated. The authors of the WEAVER++ model argued that the phonological facilitation effect might arise from a particular planning strategy adopted for dealing with the selective attention problem caused by overlapping pictures, rather than reflecting a continuous spread of activation from the distractor (which is not intended for production). To test this, Roelofs (2008) conducted three experiments with the eye-tracking techniques in which they measured not only response times but also gaze shift latencies. Gaze shifts have been consistently linked to phonological activation: if distractor pictures activate the phonological form of the nouns that designate them, they should affect not only the latency of naming the target picture but also the latency of gaze shifting. In fact, gaze shifting has been shown to be a more accurate measure of phonological length of an utterance than vocal response latencies (Levelt & Meyer, 2000; Korvorst et al., 2006). The three experiments consisted of picture-picture interference tasks (1 and 2) and picture-word interference (PWI) tasks (3, in which the distractor was alternatively a picture or a written word). In interpreting their results they considered that any phonological effect could be due to speakers mistakenly selecting the nouns of both pictures, which would force them to correct themselves. That would explain why, if phonological effects were due to an erroneous selection of the distractor picture, followed by a covert repair, the effects between phonological similar and dissimilar forms would be only present for slow responses (i.e., those in which the incorrect selection occurred). On the other hand, if the effects were indeed a reflection of cascaded processing to the lexeme, the difference in the latency between the related and the unrelated conditions would be present through most of the latency range. Moreover, in their second experiment participants named both pictures (first one of them, then the other). Hence, if phonological effects were due to the selection of the nouns of both pictures, the results of

USC UNIVERSIDADE DE SANTIAGO

naming one picture in Experiment 1 should be replicated with the naming of both pictures in Experiment 2 (in which production of both nouns is explicitly intended). Experiment 3 analysed the influence of picture distractors on word reading and vice-versa using picture-word versions of the stimuli used in Experiments 1 and 2. If effects of phonological facilitation were not obtained in this third experiment, this would mean that the amount of activation spreading from concepts to word forms was limited and task dependent. Experiment 1 showed a phonological facilitation effect across all ranges of response times, with significant differences in gaze shifting for related vs. unrelated pairs, as well as in RTs. Effects were null in Experiment 2, suggesting that in Experiment 1 the effect of phonological facilitation was not due to the erroneous selection of both pictures' lemmas for production (note that in Experiment 2 both pictures had to be selected to be produced). In Experiment 3, word distractors (which are processed firstly on the word-form level of encoding) yielded semantic and phonological effects in picture naming but picture distractors yielded no effects on word reading regardless of latency. The picture-word asymmetry is a replication of the result obtained in other studies in which the picture distractors were presented earlier to nullify the timing advantage of naming a written word (Peterson & Savoy, 1998; Roelofs, 2003). Therefore, this asymmetry suggests that the amount of activation that cascades from concepts to word forms (as seen in Experiment 1) is limited and attention dependent (Roelofs, 2003). More specifically, only when participants intentionally enhance the activation of the picture name do distractor pictures influence word reading, as observed in other studies (Roelofs, 2003). Hence, according to Roelofs and colleagues, viewing pictures passively does not lead to significant phonological activation. Consequently, they argue that the cascading of activation from lemma to lexeme is task dependent and limited, and does not normally occur in regular language production (as seen in Experiments 2 and 3).

Finally, note that the WEAVER++ model does not explain how the system is able to detect that agreement has to be fulfilled, and hence gender has to be selected. In this sense, the bypass of gender processing would only be possible if we assume that cascaded processing occurs between the lemma and lexeme stratums in normal speech within a more interactive view of lexical access. In this way, as suggested by Bordag and Pechmann (2008a), if the activation sent by gender nodes to the lexeme converges on just one morpheme or word form, feedback connections allow for activation to be sent back to the lemma, indicating that the selection of grammatical gender can be skipped.

USC UNIVERSIDADE DE SANTIAGO

4.2 INDEPENDENT NETWORK MODEL

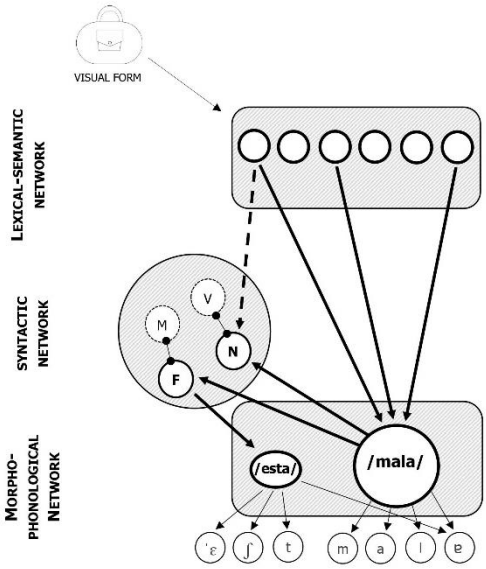
The Independent Network (IN) model is grounded mainly on data exploring the performance of abnormal populations with rare speech phenomena. Similarly to the WEAVER++ model, it distinguishes between three levels of language representation in lexical access: the lexical-semantic network, the lexical-syntactic network, and the phonological and orthographic lexeme networks. Yet the authors do not adhere to what they call the Lemma Mediation Hypothesis (Caramazza & Miozzo, 1998), which is a structural characteristic of the WEAVER++ model. WEAVER's hypothesis assumes the existence of two types of lexical nodes: the lemma and the lexeme nodes. The lemma nodes are modality-neutral and are connected to the syntactic features of words. The lexeme nodes are dependent on modality (morphemes vs. phonemes) and are connected to the morpho-phonological content of words. The IN model states that there is only one abstract lexical representation of words, one which they prefer to call lexeme rather than lemma. Hence, unlike most views of the architecture of lexical access during language production (Dell, 1990; Garret, 1992; Roelofs, 1992), in the IN model the selected lexical-semantic representation spreads activation simultaneously and independently to the lexical syntactic and the word-form networks, without a modality-neutral lemma representation as a compulsory intermediate step between semantics and word-form.

The authors highlight three reasons underlying their rejection of the Lemma Mediation Hypothesis: a) the fact that all evidence supporting the lemma as an intermediate level of lexical access is also consistent with the alternative idea of only one level of lexical representation, especially in those studies concerning the tip-of-the-tongue (TOT) phenomenon and RTs from naming studies (but for more challenging data, see Roelofs et al., 1998 and Turennout et al., 1998); b) the fact that some results with brain damaged patients are problematic in terms of the existence of modality-neutral lemma, especially those in which patients selectively produce semantic substitution errors when processing homonym words of only one grammatical class (e.g., verbs, but not nouns, such as "to watch" vs. "watch") in only one modality of output (e.g., production), while their other modality (e.g., comprehension) remains unimpaired (see Caramaza & Miozzo, 1998); (C) the fact that studies have found that the retrieval for a word's initial phoneme does not depend on the correct retrieval of the word's gender in speakers in a TOT state (speakers are no better at recalling gender information after correctly recalling the initial phoneme than when they fail), which undermines the idea that a lemma has to be fully accessed for access to phonology to occur.

USC UNIVERSIDADE DE SANTIAGO

Figure 6

Simplified structure of lexical access according to the IN model



Note. Only the Phonological-lexeme is shown, to simplify the figure. Dotted circles indicate inhibition. The lexical-semantic network does not send activation to every grammatical feature within the syntactic network. Hence, it would not activate grammatical gender, but it would weakly activate the grammatical class of the noun "mala" (and of the demonstrative "esta", but at a different moment in time, since the selection of one class entails the inhibition of the others). F = feminine; M = masculine; N = Noun; V = Verb. The flow of activation is unidirectional. Adapted from Caramazza (1997).

Hence, as can be seen in Figure 6, there are direct connections in the IN model between the lexical-semantic and the morpho-phonological strata, and both have connections to an apparently less important level of grammatical-syntactic encoding. We will now assess each of these levels in detail. The lexical-semantic network represents word meanings in a componential way; that is, as groups of semantic features. The lexical-syntactic network represents the syntactic properties of words such as grammatical category, gender, and number. More specifically, each property is organized as a subnetwork consisting of nodes representing the different syntactic functions (e.g., the masculine and feminine gender nodes). Nodes within a subnetwork compete between them and have inhibitory links. Processing ends at the phonological and orthographic lexeme networks (known as P-lexeme and O-lexeme), which consist of modality-specific representations of lexical items that are connected to the segmental information of words (i.e., the phonemes and morphemes).

USC DE SANTIAGO DE COMPOSTELA

As with the WEAVER++ model, the IN model is a forward activation model with no feedback within the production system; however, it argues that the spread of activation is cascaded. Production starts when a selected lexical-semantic representation sends activation in parallel toward the lexical-syntactic and the P- and O- lexeme networks. For instance, the semantic features of MALA such as "container", "leather", "with a handle", plus a large etcetera, are activated, which entails that the semantic representation of MALA is activated and spreads activation forward to the other strata. The semantic network only activates certain syntactic features weakly, specifically those grounded in any type of semantic information. In the case of gender, only nouns with natural gender, that is, words morphologically marked by gender, receive activation from the semantic network (Caramazza, 1997; p. 195). Thus, the gender nodes of nouns with grammatical gender receive activation exclusively from the lexeme networks. Hence, the semantic representation of a noun like "mala", whose gender value is abstract, would not activate the subnetwork of gender within the lexical-syntactic network, but will activate that of grammatical class (noun). On the contrary, the semantic representation of a noun like "menina" (female kid in Portuguese) would activate both the grammatical class and the gender subnetworks, more specifically the "noun" and "feminine" nodes. Importantly, under normal circumstances, the activation of syntactic nodes from the semantic network is not sufficient for a grammatical feature to reach the threshold for selection. Hence, selection of the full set of grammatical properties of a certain word takes place only after the activation and the selection of the modality-specific lexical node. Consequently, unlike the WEAVER++ model, the selection of the lexeme node does not depend on the prior selection of its associated syntactic features.

After the activation of this modality-specific lexical node takes place (either the O or P lexeme), activation from the selected lexeme (1) converges on the grammatical features that had already been primed by the activation coming from the lexical-semantic network, and (2) propagates to the associated phonological or orthographic segments. Hence, only after accessing the lexeme representation of /mala/ or /menina/, will their syntactic properties be selected, inhibiting the other nodes within each subnetwork (e.g., within the gender subnetwork, the masculine node will be inhibited). The selected gender node, in turn, will provide access to the phonological form of any agreement target that is intended for production. Finally, the segmental information of the noun will be activated and selected (e.g., [m] [a] [l] [ɐ]).

USC UNIVERSIDADE DE SANTIAGO

To conclude, of particular note is that the IN model does not take a stand in terms of whether or not a context of agreement is necessary for the gender of a noun to be selected. However, in light of the WEAVER++ idea, the authors do accept that this may be possible (Caramazza & Miozzo, 1998; p. 238). If this were the case, the IN model would accommodate such view, as it proposed: (1) a direct connection between the conceptual and the word-from networks without the mediation of the syntactic network, (2) a total lack of dependency between the phonological and syntactic networks, so that the selection of a lexeme does not guarantee the selection of the full set of its associated syntactic features (Caramazza, 1997; p. 197).

In Table 4, we outline the main characteristics of both models.

Table 4

Summary of the main characteristics of the WEAVER++ and IN models

	WEAVER++	**IN model**
Type of data basis	Reaction times of healthy population	Performance of brain damaged patients
Computational implementation	Yes	No
Semantic representations	Non-Componential	Componential
Lemma Mediation Hypothesis	Assumes	Rejects
Cascading to word-form stratum	Weak and Task limited*	Existent and similar for all the levels of processing
Spreading of activation	Forward (bidirectional between conceptual and lemma strata; unidirectional to lexeme but bidirectional for monitoring*)	Forward (unidirectional)
Selection of nodes within a stratum	Competitive	Competitive
Selection of syntactic nodes within the syntactic stratum	Competitive	Competitive with inhibitory links
Distinction between activation and selection	Yes	No
Presence of agreement for gender selection	Yes	No, but possible

Note. Characteristics marked with an asterisk were only included in the most recent versions of the model.

5 EXPLORING GRAMMATICAL GENDER IN LANGUAGE PRODUCTION

As seen in the second chapter of this thesis, grammatical gender is usually defined as an abstract lexical-syntactic property of nouns (Schriefers and Jescheniak, 1999). From this definition, it becomes clear that gender has two components: a lexical component related to nouns, and a syntactic component related to word forming. On the one hand, grammatical gender is a property of nouns, since they are the units that store or allow the computation of any gender value (van Berkum, 1996). On the other hand, grammatical gender is a syntactic property of language, critical to establishing proper communication in that it serves the function of determining the form of many of the words contained in speech. Due to the central role that gender plays in the structure of language, its assessment has focused mainly on its syntactic component through the analysis of agreement relations (see Deutsch et al., 1999; Friederici & Jacobsen, 1999). However, as grammatical gender cannot be reduced to its agreement system (Colé et al., 2003), the study of its lexical component as related to nouns themselves is also crucial for the proper understanding of its representation and processing in the mind. This is the main aim of the present thesis.

The study of the representation and processing of grammatical gender during the lexical access of nouns has led to many theoretical disagreements in the recent decades. Several debates remain unsolved, including those relating to: a) the role of agreement in gender selection; b) the automatic vs. competitive-based nature of the mechanisms underlying gender selection; c) the cross-linguistic differences that gender representation and processing might have; and d) the role of gender transparency in language production. More specifically, it is unclear what the role of agreement is in the selection of gender, and different authors make differing claims. For some of them an agreement context is compulsory for the gender of a noun to be retrieved (La Heij et al., 1998; Schriefers, 1993); for others, gender seems to be retrieved regardless of the presence of an agreement context (Cubelli et al., 2005). Until the first decade of the 21st century,

the majority of studies, which were mostly conducted with Germanic languages, showed that agreement contexts were actually the direct source of any gender-related effect we could find, in that the gender of nouns is retrieved through automatic mechanisms that do not entail activation (Caramazza et al., 2001). In the absence of different degrees of activation for each gender value (e.g., masculine vs. feminine), there is no competition for selection, and competitive effects of gender cannot be obtained: only effects of agreement, understood as competition between different agreement forms (e.g., definite articles of different gender) can be studied. More recent studies with Romance languages, however, showed competitive gender effects without agreement being required (Cubelli et al., 2005; see Finocchiaro et al., 2011 for null results). The role of transparency is equally unclear. As seen in Chapter 4, both phonological gender transparency (the presence of grammatical gender regularities in nominal endings) and the general degree of gender transparency of a language have repercussions for language acquisition and comprehension. The influence of these factors in language production is unknown.

In what follows we will address these issues by conducting a detailed examination of those studies available to date in both the monolingual and the bilingual domains.

5. 1 EVIDENCE ON NATIVE LANGUAGE PRODUCTION

The representation and processing of grammatical gender during native language production is assessed through a variant of the Stroop task: the PWI paradigm (Klein 1964; Rosinski et al., 1975). Within this paradigm, participants are asked to name aloud pictures that are presented together with superimposed written distractor nouns using bare nouns or noun phrases. The target nouns that designate the pictures and the distractor nouns are known to compete for selection. This is reflected in two typical effects explored through this task: the semantic interference and the phonological facilitation effects (Caramazza & Costa, 2000; Lupker, 1979, 1982). When exploring grammatical gender during lexical access, modulations in the response times are expected depending on the congruence between the gender values of target and distractor nouns. More specifically, what is theoretically expected is the so-called "gender congruency effect" (GCE, Miozzo et al., 2002), by which participants are faster when naming targets paired with distractors of the same gender (for a review, see Wang & Schiller, 2019).

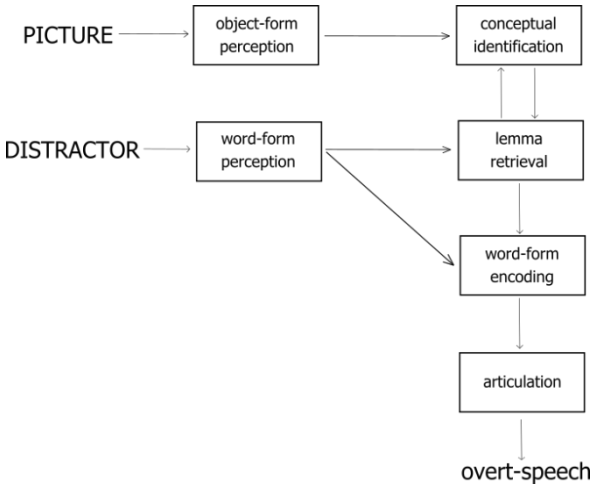
The most detailed predictions regarding gender competition come from the WEAVER++ model. Since it is grounded on measures of RTs and is heavily based on studies featuring the

USC UNIVERSIDADE DE SANTIAGO

PWI paradigm, the authors made a specific assessment of the mechanisms underlying this task. Hence, in Figure 7, we can see what is expected to occur within the paradigm during lexical access (Roelofs, 1992). Whereas the picture has direct access to the related conceptual information, the distractor word is thought to have direct access to both the lemma and the word-form level of encoding, as supported by the literature on reading and word recognition (e.g., Altmann & Davidson, 2001; Roelofs, 2003).

Figure 7

Stages of lexical access in the PWI paradigm as conceived by the WEAVER++ model



Note. Adapted from Roelofs (1992).

The target would activate its own gender node at the lemma level, and the same would happen with the distractor and its respective gender. Whenever there is a context of agreement in the response to be given by the participant, gender is necessary for fulfilling agreement, and hence, gender nodes will not only be activated but also selected. If selection is necessary, then competition for selection occurs between nodes. As described in the previous chapter, the first step for node selection in WEAVER++ is reaching a critical difference in the degree of activation between the target node (that of the noun to be produced) and all the other competitor nodes. Hence, the existence of an active competitor that is being boosted through a written word form (i.e., the distractor) slows down the selection of a target node since it takes more time to reach a critical difference between the degree of activation of the target node and that of the boosted competitor. In the case of grammatical gender, the target's gender node (e.g., in Portuguese, the feminine node for target "*mala*" [purse]) would either reach its threshold for selection faster thanks to the activation received by a distractor of the same gender (e.g., a

USC DE SANTIAGO DE COMPOSTELA

feminine noun such as "*seta*" [arrow]), or this process would be hampered because of the activation of the other gender node by a gender incongruent distractor (e.g., a masculine noun such as "*bolo*" [cake]).

Note that the fact that the distractor has direct access to the lemma stratum should induce an earlier activation of its gender node in comparison to that of the target's gender node, which because the target has first to be encoded at the conceptual stratum. Thus, if target and distractor are presented at the same time, a time gap should exist between the activation sent to gender nodes by both stimuli. To observe gender effects with a simultaneous presentation of both target and distractor, the gender node activated by the distractor should remain active for a certain period, allowing both flows of activation to coincide. However, naming and lexical decision studies on gender processing combined with the paradigm of priming (i.e., targets are preceded by a briefly presented prime which can be related or not with the target) have shown that gender priming effects are difficult to obtain, even when considering priming between elements of agreement (e.g., definite articles and nouns; Carello et al., 1988; Jescheniak, 1999; Schmidt, 1986; van Berkum, 1996, 1997). This may be due to a rapid decay of the level of activation of the gender nodes. Hence, the WEAVER++ model should predict effects (or at least, greater effects of gender) when the distractor is presented later than the target (i.e., when a positive Stimulus Onset Asynchrony [SOA] is used).

The authors of the IN model did not explicitly assess competition between gender values within a PWI paradigm in their original proposal. However, there are two important points regarding the retrieval of grammatical gender in the IN model: (1) the abstract phonological representation of the word receives activation earlier than the features at the syntactic network (including grammatical gender); and (2) the nodes at the syntactic network are thought to receive activation and compete, and within the same subnetwork the node that is selected inhibits the other nodes. Caramazza also claims that syntactic nodes reach a "threshold" for selection (Caramazza, 1997; p. 195), but no further information is specified. Thus, for instance, a target noun such as "mala" (purse) would activate the feminine gender node, and a distractor such as "bolo" (cake) would activate the masculine gender node. The feminine gender node will reach the threshold for selection because it is the target noun to be produced, but it will compete with the masculine gender node activated by the distractor. This competition should hypothetically hamper the selection of the feminine node. Hence, the original IN model would predict a GCE by which target-distractor pairs of different gender will slow down the response

USC UNIVERSIDADE DE SANTIAGO

but those of the same gender would make it easier to reach the threshold for selection of the target gender node, thus speeding up the response. Importantly, as happens in WEAVER++, the architecture of lexical access proposed in the IN model suggests that the gender node of the distractor would be accessed earlier than that of the target. Again, this delay would predict stronger gender effects with a positive SOA for the distractor.

As mentioned above, the IN model makes no predictions relating to the retrieval of gender depending on the presence of an agreement context. In any case, the architecture proposed by the model allows for the bypassing of gender selection through the direct link between the semantic and the lexeme networks. In this sense, if the task is to name a picture as fast as possible, it may be the case that gender is only selected if a context of agreement is necessary, similarly to what is proposed by the WEAVER++ model.

The predictions based on these models have promoted a debate as to what kind of phrases participants should use to name the pictures. If we are willing to study the representation and processing of grammatical gender as a feature directed linked to nouns, but a context of agreement is compulsory for the selection of grammatical gender to occur, then it is only logical to ask participants to name the pictures using nouns plus other elements of agreement; this because, theoretically, if there is no context of agreement, gender nodes may not compete for selection or even be active, and hence the response times of the participants will not be affected in terms of gender by the presence of any other nouns. Indeed, the first author to use the PWI paradigm within the study of grammatical gender selection was Schriefers (1993), who tested native speakers of Dutch and asked them to use noun phrases rather than bare nouns. These noun phrases consisted of either a definite article, plus an adjective, plus a noun (e.g., “de groene tafel” [the green table]), or an adjective plus a noun (e.g., “groene tafel” [green table]). The definite article varied according to gender (e.g., “de groene tafel” [common gender] vs. “het groene huis” [neuter gender; “the green house”]), and so did the adjective when presented alone (e.g., “groene tafel” vs. “groen huis”). Target and distractor were presented at the same time. The results showed what would become the classical GCE: participants were faster when targets and distractors had the same gender in comparison to when their gender differed. Schriefers (1993) located the effect at the level of gender feature selection, and interpreted it following the tenets of the WEAVER++ model: the gender values of the target and the distractor were represented at the lemma level and could activate the same or a different gender node, facilitating or hampering the retrieval of gender. This interpretation was soon called into

USC UNIVERSIDADE DE SANTIAGO

question and an alternative hypothesis was advanced: the GCE in fact reflected competition between determiner forms (de vs. het) rather than between abstract gender values (common vs. neuter) (Miozo & Caramazza, 1999). Schiller and Caramazza (2003) hence distinguished between the Gender Selection Interference Hypothesis (GSIH) and the Determiner Selection Interference Hypothesis (DSIH). Due to cross-linguistic differences, literature regarding both hypotheses has to be considered according to the language family.

5.1.1 Studies with Germanic and Slavic languages

The GCE has been systematically replicated in German when considering noun phrases formed by either definite articles, adjectives, or demonstratives plus nouns (Bürki et al., 2016; Heim, Friederici, et al., 2009; Schiller & Caramazza, 2003; Schiller & Costa, 2006; Schriefers & Teruel, 2000), Dutch (La Heij et al., 1998; Schiller & Caramazza, 2003, 2006; Starreveld & La Heij, 2004; van Berkum, 1997), and Czech (Bordag & Pechmann, 2008), as well as when considering referential processing through pronouns in Croatian, a language which does not have articles (Costa, Kovacic, Fedorenko, et al., 2003).

Importantly, the effect is absent if: (1) bare nouns are used to name the pictures (e.g., "*tafel*"; Finocchiaro et al., 2011; La Heij et al, 1998; Starreveld & La Heij, 2004), or (2) the form of the determiners does not vary across genders (e.g., Schiller & Caramazza, 2003, 2006). In an influential study, Schiller and Caramazza (2003) tested the following hypothesis: if the GCE indeed reflects competition at the level of grammatical encoding, it should not matter whether or not the form of the definite article varies across gender values. Hence, they asked both German and Dutch participants to name pictures either in the singular (in which the articles "*de*" [Common, C] and "*het*" [N] are used in Dutch, and the articles "*der*" [M], "*die*" [F], "*das*" [N] are used in German) or in the plural (in which solely the articles "*de*" in Dutch and "*die*" in German are used regardless of the gender value of the head noun). The results were very clear: the effect was only obtained when the article varied across gender values (the singular condition). In another study, Schiller and Caramazza (2006) arrived at a similar conclusion. In Dutch, all diminutive forms take the suffix "*-je*", and by taking this suffix the noun is preceded by the article "*het*" regardless of its gender value in its standard form (e.g., "the book" translated as "*de boek*" transforms into "the little book" as "*het boekje*"). The results consistently showed a GCE in the reverse direction when common gender nouns were used in their diminutive form, which ultimately supports the notion that the RTs reflect competition between articles. More

USC UNIVERSIDADE DE SANTIAGO

specifically, common target nouns in their diminutive form use the article “het”, and thus they are gender incongruent with a neuter distractor, but determiner congruent in that both are preceded by the article “het”. Therefore, in the diminutive target condition, when the target was common and the distractor was neuter, they were preceded by the same article (het) and facilitation for the condition of gender incongruency was obtained. For the standard form of common and neuter targets, as well as diminutive neuter targets, the classic GCE was observed. Finally, Bordag and Pechmann (2008a) found similar results in Czech: the GCE was absent when ordinals that did not vary across gender values were the only elements preceding the nouns.

Overall, the DSIH seems to be a better fit for the available evidence on the Germanic and Slavic languages tested thus far. Nevertheless, neither WEAVER++ nor the IN model can accommodate the DSIH within the current state of their tenets. Initially, the fact that Schiller and Caramazza (2003, 2006) and Bordag and Pechmann (2008a) found null effects with nouns in their plural, diminutive, and invariant forms led Levelt et al. (1999) to try to explain them via their original formulation of WEAVER++. According to the authors of this model, grammatical gender is only selected when required for agreement, that is, when required to determine the form of the other words in speech. Only in the presence of agreement can competition for selection occur. As previously noted, the mechanisms for how this checking procedure occurs have not been theoretically or experimentally uncovered, although they would require the WEAVER++ model to accept the existence of bidirectional connections between lemma and lexeme within the production system. Despite the fact that this mechanism has not been detailed, if we assume that agreement is indeed necessary for gender to be selected, then, since plural definite articles in German and Dutch, articles for diminutive forms in Dutch, and certain ordinals in Czech, do not vary across gender values, it could be claimed that gender was not required and hence, not selected. If this conjecture were correct, the GSIH would still be a valid proposal and the results concerning plural and diminutive forms could not be taken as supporting evidence for the DSIH.

However, there is a body of literature based on what has been called the singular-plural paradigm (Jescheniak et al., 2014) that clearly shows that the nouns’ gender feature is indeed considered in the production of plural and diminutive noun phrases. Regarding number, it seems that the singular is selected by default, and hence the singular definite articles for every gender value compete for selection along with the plural definite article. More specifically, Janssen

USC UNIVERSIDADE DE SANTIAGO

and Caramazza (2003) used simple picture naming tasks with noun phrases consisting of singular or plural definite articles plus nouns in Dutch. As explained above, in this language, in the singular, the definite article for the common gender value is “de”, whereas for the neuter gender value it is “het”, but in the plural “de” is the definite article for both values. The authors compared the cost of producing a neuter noun in the plural to that of a common noun in the plural. If gender features were being selected and singular determiners activated even though the plural was the form to be produced, then there should be a cost for neuter nouns (“het” [singular] vs. “de” [plural]) in comparison to common nouns (“de” and “de”). Indeed, longer naming latencies were consistently found when the singular and plural determiners differed than when they were the same (i.e., a cost for the neuter value). Schriefers and colleagues (2002), as well as Schriefers et al. (2005), replicated the effect in German (i.e., the cost for the masculine and neuter [definite articles “der” and “das”] in comparison to the feminine, whose definite article coincides with the one for plural [“die”]). Janssen and Caramazza (2003) also showed similar results when considering standard vs. diminutive forms. More specifically, diminutive noun phrase naming was faster for neuter gender nouns (in which the definite article “het” coincides with the article for diminutive forms of any gender) than for common nouns (in which the definite article does not coincide, since it is “de” [vs. “het”]). This effect was also replicated by Spalek and Schriefers (2005).

All in all, the studies mentioned above show that the gender feature contributes to the selection of the determiner even though it is logically unnecessary: the singular form in the case of the plural is always considered, as well as the standard form in the case of diminutives. These, then, are “default” forms (see the singular-as-default hypothesis, Janssen & Caramazza, 2003). In sum, gender is being activated in the case of plural and diminutive noun phrases, and hence the explanation in WEAVER++, whereby the GCE is absent because gender is not being selected due to the absence of agreement in the plural and diminutive forms, is not valid. It is not currently known whether gender is actually being selected or not in Czech for gender invariable ordinal forms.

Since gender is indeed being selected but the competitive effects are reflecting the selection of determiner forms, Schiller and Caramazza (2003, 2006) interpret their results as follows: the GCE is a determiner congruency effect, and because the determiner forms are the ones competing, gender is a feature that is automatically retrieved because of lexical selection, and hence does not involve gender values competition. This claim can be incorporated into both the

USC UNIVERSIDADE DE SANTIAGO

WEAVER++ and IN models, but some modifications must be made. On the one hand, according to Schiller and Caramazza (2003), if the account of the DSIH is right, it would prove the existence of cascaded processing from the distractor to the articles' phonological forms. Hence, WEAVER++ should accommodate a cascaded spread of activation. More precisely, the DSIH implies that the determiner form of the distractor is activated even though the distractor lexical node itself is not selected for production, which confirms the existence of activation cascading from an active-but-not-selected noun lemma to the lexeme of its potential determiners (Schiller & Caramazza, 2003; p. 966). However, we are unsure about this rationale since it is expectable that the distractor lexical node (lemma) is indeed selected because it has been read. Once selected, due to language comprehension, activation spreads to the gender nodes and consequently to the definite articles, which are intended to be produced when naming the target. Hence, for us it does not seem clear that the DSIH forces the WEAVER++ model to accommodate cascading processing (note however that the latest updates of the model do consider the existence of a limited cascaded processing). We think that the challenge here is simply to change the basic tenets of the model regarding how grammatical features are selected: grammatical features (at least gender) would be accumulating different degrees of activation but not competing. Schiller and Caramazza (2003, 2006) argue that the DSIH and the idea of gender as an automatically selected feature are "naturally" assumable within the IN model. However, we believe that this is not entirely correct. According to this model, gender nodes within the "gender subnetwork" do accumulate activation, reach a "threshold for selection", and "have inhibitory links since they are in competition" (Caramazza, 1997, p. 194). The model would have hence to change its statements to conceive that the selection of a node gathering activation within a subnetwork involves the inhibition of the others without competing with them, which may appear as a strange idea.

We will analyse in more detail the reliability of these hypotheses (GSIH vs. DSIH), taking into account all available corpus evidence from the literature on Romance languages. Before doing so, we will explore three final issues concerning Germanic and Slavic languages: the results of studies with brain activity measures, the GCE as a function of the SOA, and the type of agreement element (freestanding vs. bound morphemes).

USC UNIVERSIDADE DE SANTIAGO

Two additional studies with the PWI paradigm used direct measures of the brain activity to explore the GCE. The study by Bürki et al. (2016) tested native speakers of German while recording their brain waves through EEG measures. Participants were asked to exclusively use

noun phrases (definite article plus noun) to name the pictures. The targets and the distractors varied in another factor besides gender congruency: phonological overlap. In this way, the authors had a condition that could reflect word-form encoding and provide a time window to study any other effect related to gender congruency. They argued that gender retrieval would supposedly occur before noun-form encoding. Therefore, modulations in the brain waves related to the gender congruency conditions observed before a phonological facilitation effect would support the GSIH; the opposite would apparently support the DSIH. RTs revealed a classical GCE, plus the phonological facilitation effect. Response-locked event-related potentials (ERPs) were analysed and amplitudes differed as a result of phonological overlap in several time windows starting from 480 ms before the onset of articulation. The difference in the amplitudes between the gender congruent and incongruent conditions did not overlap with the phonological facilitation effect and started 210 ms before the onset of articulation, lasting for 60 ms. The phonological overlap conditions still showed an effect in a late time window after the GCE had taken place: 165 to 130 before the onset of articulation. Since the GCE occurred after the processing of the phonological form of the noun had started, Bürki and colleagues interpret this as support for the DSIH. The authors also state that the fact that determiner selection occurs after word form encoding is not surprising, and that even if the form of the determiner does not depend on the form of the noun, determiner form processing takes place after a noun's form processing. Unfortunately, they do not explore the differences on the localization of each brain modulation depending on the effect.

We think that the interpretation of the data, however, is far from clear. An interactive processing or a parallel approach of the WEAVER++ model, as conceived by Dell's model (1986, 1990), would not predict such rigorous sequentiality between gender processing and word-form processing. In fact, the data from Bürki and colleagues data show that the "gender" congruency effect occurs "during" the phonological facilitation effect, even though they do not interact, and this is compatible with a parallel processing during lexical access. The phonological overlap conditions were, in fact, the last to show an effect. The authors are unsure about the interpretation of this late effect and suggest possible monitoring processes. Anyhow, the IN model, which has been strongly supported by those authors who also support the DSIH, states that the phonological form of the noun (the lexeme) should be processed before gender, and the morpho-phonological segments (which receive activation from the lexeme) may be processed in parallel with the syntactic characteristics of the noun and perhaps the associated

USC UNIVERSIDADE DE SANTIAGO

determiners. This view would not allow us to discriminate between the GSIH and the DSIH according to the results of this specific study. Besides, it seems unclear to us what the time course of lexical access is in a language like German according to Bürki and colleagues and their view based on the WEAVER++ model. It appears that gender would be processed before the phonological form of the noun, but activation would somehow not be spread to the respective determiner form and would wait for the phonological form of the noun to be encoded, even though the phonological form of the noun is not at all necessary to establish the form of any element of agreement in the language. We are uncertain about how much sense this rationale makes. Finally, Bürki and colleagues point out a troubling issue regarding the discrimination between target and distractor processing in the ERP effects. Indeed, the morpho-phonological information of the distractor is processed before that of the target, and early effects of phonological overlap such as those we have seen may be affected by this fact. However, due especially to the tenets of the IN model, according to which phonological effects would occur earlier and facilitation by gender and by determiner selection would occur afterward, we believe that an optimum approach here would have been to conduct an experiment in which the phonological overlap and gender congruency conditions were maintained but bare nouns were used to name the pictures. To sum up, it seems to us that the conclusions drawn from this study are tentative, and that more research is clearly needed.[11]

There is another experiment featuring a PWI task (combined with the fMRI technique) that merits attention here. Heim, Friederici, and colleagues (2009) tested native speakers of German with the same materials used in Schiller and Caramazza's study (2003). Participants were asked to name the pictures using noun phrases (definite article plus noun). They found a classical GCE on the RTs, but critically, they also found that the haemodynamic response (the activation) on the left BA 44 increased faster when the picture name and the distractor had the same gender. Previous literature shows that language production in this area is related to the retrieval of bound morphemes. In their interpretation of the results, the authors first say that they provide evidence that this area is also related to the retrieval of freestanding forms, since they argue that

[11] We are aware of Wang et al.'s (2019) study with classifiers and EEG measures in Mandarin Chinese. The authors see other classifiers as potential holders of the same type of congruency effects as grammatical gender. However, we are not sure that other classifiers can be compared to grammatical gender. As we have seen in Chapter 1, classifiers are independent items attached to words (such as nouns) and are mostly semantic driven while grammatical gender does not entail the attachment of items and it is purely abstract and *grammaticalized* (similarly to inflectional classes). Classifiers in Mandarin form more than a hundred classes; they are not part of bipartite or tripartite systems. Moreover, as seen, there are different types of classifiers such as numeral, noun, or relational classifiers, and the authors do not specify and discuss what type of classifiers are being put into analysis and equalized to grammatical gender in Mandarin. We thus opted to leave these results aside, as they may be potentially confounding for the debates here addressed.

the GCE is a determiner congruency effect. Crucially, they end up revealing a problem with their data: neuroimaging studies suggest that there is a more fine-grained functional segregation of the left BA 44 according to which their interpretation does not fit the current evidence. Whereas the retrieval of bound morphemes and phonological codes has been shown to involve the superior portion of the BA 44 (e.g., Heim et al, 2003, 2005; Longoni et al., 2005; Marangolo et al., 2006; Miceli et al., 2002), syntactic processing of features such as grammatical gender or word class seems to show up in its inferior portion (e.g., Heim et al., 2003, 2005; Longoni et al., 2005). However, the effects that they obtained were restricted to the inferior portion of the left BA 44, and did not show up in the superior portion, in which phonology and bound morphemes are encoded. The authors admit that an explanation for such discrepancy might be that the effect they found on the fMRI reflects gender selection that occurs prior to the selection of the determiner form. Importantly, they also indicate that the division on the left BA 44 according to the type of information that is being encoded is established for language comprehension, but unclear for language production. In the same year, Heim, Eickhoff, et al. (2009) conducted a naming task with priming, in which prime and target varied as a function of semantic relation, gender congruency, and phonological overlap. Again, participants were native speakers of German. Critically, they were asked to name the pictures using bare nouns. Although the results concerning RTs were not clear enough, the fMRI revealed many highly significant effects, including that the same area of the BA 44 showed modulations regarding the gender congruency of prime and targets, which differed from the effects of phonological overlap. This ultimately supports the GSIH and the interpretation of Heim, Friederici's, et al. (2009) congruency effect as a gender effect, and not a determiner effect.

Turning to the role of the SOA in the GCE, we reasoned at the beginning of this chapter that both the WEAVER++ model and the IN model would expect either (1) a GCE only with positive SOAs, or (2) a greater GCE with positive SOAs. However, the available evidence nullifies the first prediction. It does not allow us, however, to confirm or deny the second prediction. The simultaneous presentation of both target and distractor (SOA = 0 ms) has been shown to be a satisfactory choice (Bürki et al., 2016; Costa, Kovacic, Fedorenko, et al., 2003; Heim, Friederici, et al., 2009; La Heij et al., 1998; Schriefers, 1993; Schiller & Caramazza, 2003, 2006; Starreveld & La Heij, 2004; van Berkum, 1997). However, the effect has been also obtained in Dutch for negative SOAs of -200 ms and -100 ms (when distractors are presented before the targets), and with positive SOAs of +100 ms, with no interaction between gender

USC UNIVERSIDADE DE SANTIAGO

congruency and SOA as factors (Schiller & Caramazza, 2003; Schriefers, 1993). From the perspective of both the GSIH and the DSIH (in which the access to the determiners depends on the previous selection of gender features), we may think that the results with SOA = 0 suggest some kind of spreading of activation going from the word-form to the syntactic strata. This would compensate the delay suffered by the target in the activation of gender in comparison to that of the distractor. The WEAVER++ model would hence require bidirectional connections to send activation from the lexeme to the lemma, thus reactivating the gender nodes (that had already been activated via the distractor by the direct connection of the perceptual information to the lemma). In this sense, an interactive view of lexical access as proposed by Dell's model (1986, 1990) would accommodate these results. It is true, however, that the newer updates of the WEAVER++ model (Roelofs, 2004) carefully describe a feedback of activation that can occur through the monitoring system of the internal word-form representations. More specifically, the activation of word-form information of the distractor at the lexeme would be verified by sending back signals to the perceptual system to re-analyse the distractor word, which in turn would activate the lemma again (and the gender nodes). It is unclear, however, if such a process featuring these multiples steps would affect the gender selection of the target noun, which perhaps would already have been selected before the distractor's gender information had gained a reactivation through the monitoring system.

From the perspective of the IN model, it seems harder to accommodate the results: the distractor would activate directly the orthographic segments plus the abstract representation of the word at the word-form level (i.e., the lexeme). In turn, the latter would activate gender and its respective determiner form. Nevertheless, the target noun would first be encoded in the semantic network, then at the lexeme level, and finally it would activate gender in the syntactic network and consequently its respective determiner form. Since the activation coming from both stimuli start from different points, a possible explanation for a re-activation or maintained activation of a distractor noun would be the existence of bidirectional connections between orthographic segments and the lexeme, in line with highly interactive models of language production (see Dell, 1990). In any case, there may be a simpler explanation related to the task itself: the distractor, in comparison, for instance, to a prime word, remains on the screen. The continuous availability of the distractor may allow the activation of the gender nodes to be maintained and their decay rate to be reduced. In this sense, Schriefers and Teruel (2000) tested German speakers and failed to obtain a GCE with a SOA of 0 ms, but systematically obtained

USC UNIVERSIDADE DE SANTIAGO

it with positive SOAs (+75 and +150 ms). Critically, in their study the presentation of the distractors was auditory: the unavailability of the distractor to the perceptual system of participants may have ceased the flow of activation that a written distractor noun is able to maintain through continuous exposure. Hence, auditory word stimuli, like written stimuli, seem to directly access the lemma (or the lexeme in the IN model), and this time-based advantage compared to a picture that is first processed at the conceptual stratum (or semantic network) may entail the unmatched activation of gender nodes (and the determiner forms) by distractor and target, creating the need for a positive SOA for the effects to be observed.

Another debate that seems to generate more questions than answers is that of the type of agreement element that may be behind the "GCE" (i.e., the determiner congruency effect)[12]. On the one hand, Schriefers (1993) obtained a GCE in Dutch with adjective-noun phrases, although it was smaller than that for noun phrases formed not only by adjectives but also by articles (31 ms vs. 56 ms). Importantly, while adjectives vary across gender through bound morphemes attached to the stem of the word (e.g., in Dutch "groen" [N] vs. "groen**e**" [C]), definite articles have different whole forms per gender, so they are referred as "freestanding" forms (e.g., "het" vs. "de"). Bordag and Pechmann (2008a) conducted a PWI study with native speakers of Czech, a Slavic language. Czech has three genders (masculine, feminine, and neuter) and two types of adjectival declension: hard and soft. In the nominative case, in the singular, each gender has a different ending in hard declension, but one invariant ending for the soft declension. Ordinals adhere to this system, and thus some ordinals have hard and others soft declensions. Importantly, ordinals with hard declensions modify their form through bound morphemes since gender is revealed through inflectional suffixes attached to the stem. In line with Schriefers (1993), the results showed a GCE when participants were asked to name pictures with noun phrases consisting of either demonstratives (freestanding morphemes, 26ms) or ordinals with hard declensions (bound morphemes, 37ms). However, Schiller & Caramazza (2003) observed the contrary. In Dutch and German, when noun phrases were formed by adjectives rather than freestanding phonological forms (definite articles), the GCE was absent. Costa, Kovacic, Fedorenko, et al. (2003) obtained similar results in Croatian when testing gender-marked personal pronouns, which are freestanding forms, and gender-marked possessive pronouns, which are inflected through bound morphemes. The GCE was only found for the former in SOA 0 (the only SOA tested) and was absent for the latter in different SOAs

[12] The use of the quoted acronym GCE will be henceforth used to refer to the determiner congruency effect.

(0, +75, +150). Schiller & Costa (2006) extended these results to other closed-class words. They failed to find effects in German using indefinite articles in masculine and feminine, which are marked through bound morphemes (ein [M] vs. ein**e** [F]), but systematically obtained the GCE with definite articles (der [M] vs. die [F]). These results show that the effect does not depend on the grammatical class of the gender-marked element and provides evidence that the key is somehow in the opposition between freestanding elements and bound morphemes. Schiller and colleagues restricted the DSIH to freestanding phonological forms by stating that, indeed, grammatical features automatically become available as part of the lexical node selection process, but if the phonological realization of grammatical features results in different lexical forms, there is interference due to competition at the level of form selection. In the case of bound morphemes, against the views of other authors, such as Garret (1982), but following the ideas of Lapointe and Dell (1989), Schiller and colleagues claimed that they are not subject to competition since they involve subsequent "phonological transformations" of the base-stem rather than the selection of independent morphemes by competition (Lemhöfer et al., 2006).

The literature, however, is far from clear and Schiller and colleagues seem to be establishing conclusions that are too strong for an area in which two studies find an effect whereas another three fail to find it. Moreover, the evidence on simple picture naming tasks within the singular-plural paradigm shows that bound morphemes are selected in a similar manner to freestanding forms (i.e., by competition). Schriefers et al. (2005) tested German speakers within this paradigm and considered two types of noun phrases, those with definite articles, and those with adjectives. For both types of elements of agreement, the feminine has the same form or inflective morpheme for the singular and the plural ("die" and "-e"). The results showed a significant advantage for the feminine gender value in comparison to the masculine and the neuter, this for both definite articles and adjectives in the plural, although it was smaller for the latter. Lemhöfer et al. (2006) replicated Schriefers et al.'s experiments in Dutch and obtained similar evidence. Importantly, although there was an attenuation of the effect with bound morphemes, it was not significant. These studies provide strong evidence against the idea that freestanding and bound morphemes are processed differently. The reason behind the null results obtained by Schiller et al. remains unknown. What seems clear is that even though the competition effects for bound morphemes may be harder to capture, they are indeed obtainable. When obtained, except in the case of Bordag and Pechmann's study, they tend to be smaller than those for freestanding elements. Schriefers et al. (2005) hypothesized

USC UNIVERSIDADE DE SANTIAGO

that competition processes might be more difficult to detect behaviourally by measuring onset naming latencies for elements that occur in non-initial positions of the utterance than for elements that are in the initial position. Indeed, bound morphemes tend to appear later in the utterance (e.g., "**die**" vs. "ein**e**"). Costa, Kovacic, Fedorenko, et al. (2003) argued that in their study, gender-marked bound morphemes appeared in an earlier position ("moj clarinet" [M] vs. "moj**a** truba" [F]; "my clarinet/trumpet") than the gender-marked freestanding morphemes did (i.e., personal pronouns after a verb, "vidim ga" [M] vs. "vidim je" [F], "see it"). The GCE was only present for the freestanding elements (the personal pronouns). Based on this, Costa and colleagues conclude that the explanation by Schriefers grounded on the time course of lexical access is invalid. However, we consider this response to Schriefers and colleagues' hypothesis to be not entirely correct. First, it is unclear that the comparison of the time course of lexical access in regard of gender is the same for noun phrases and verbal phrases with elided referents in which pronouns are used as the gender marked elements. In noun phrases, the encoding of the noun lemma or lexeme provides access to gender and the respective phonological freestanding forms of various elements; for verbal phrases, the process may be similar, but it entails the inhibition of the noun, perhaps requiring an extra amount of cognitive resources and additional time for the response to be ready for articulation. In the case of noun-adjective phrases or noun phrases formed by indefinite articles (ein/e) and nouns, the participant does not have to inhibit any noun or produce a word class with no gender (the verb). In these cases, the stem is always the same (e.g., "ein-"), and this can indeed produce some kind of task-dependent strategy or tendency to initiate the answer as soon as possible, relying on the use of the same stem. Importantly, this is in line with the following issues: (1) an inspection of Costa, Kovacic, Fedorenko, et al.'s (2003) study shows that RTs tend to be higher for their experiment with verbal phrases in comparison to their other two experiments with noun phrases formed by possessive adjectives with bound morphemes (712.5 ms vs. 680 ms vs. 669 ms; but note that those were different participants); (2) an observation made by Jescheniak et al. (2014) that the greater the size of the response set of words with gender-inflected bound morphemes (e.g., 5 possible colour adjectives vs. only 1), the higher the chance of observing competition between bound morphemes (i.e., the participant cannot develop a task-dependent strategy to initiate the response with the same stem). In the studies by Schriefers (1993) and Bordag and Pechmann (2008a), participants were provided with a large response set of four adjectives and two ordinals, respectively (i.e., four stems plus two different suffixes, and two stems plus with three

USC UNIVERSIDADE DE SANTIAGO

different suffixes, respectively). Studies by Costa, Kovacic, Fedorenko, et al. (2003) and Schiller and Costa (2006) only provided one element or stem (Croatian possessive, "moj-"; German indefinite article, "ein-") with two different suffixes in the response set. The inclusion of different stems in the response set impedes the participant from developing a task-dependent strategy to start the response as fast as possible using the same initial element. As stated by Schiller and Caramazza (2003):

> The use of a unique stem may allow participants to prepare for the production of the stem and begin speaking as soon as they have enough information about the noun, thereby masking any effects of competition that may occur in the selection of the gender-marked suffix. (p. 178).

In our view, this would reduce the RTs of the participants and make the onset of the response insensitive (or less sensitive, at least behaviourally) to the interference occurring between elements of different gender. To test the validity of this hypothesis, Jescheniak et al. (2014) replicated Lemhöfer et al.'s (2006) singular-plural paradigm study with a smaller response set. We recall that in the singular-plural paradigm, bound morphemes have consistently shown competitive effects, and so it should be easier to obtain them than in the PWI paradigm. Hence, instead of the two adjectives used by Lemhöfer et al. (2006), Jescheniak and colleagues used only one, the same number of stems used by Costa, Kovacic, Fedorenko, et al. (2003) and Schiller and Costa (2006) for their bound morphemes. Indeed, they failed to replicate the cost previously seen for neuter nouns in the plural. Interestingly, whereas the RTs in Lemhöfer et al. (2006) with various stems were of a mean of 763 ms, for Jescheniak et al. (2014) they were of 684.5 ms, which may indicate that this task-dependent strategy speeds up the onset of the response. In conclusion, the probability of observing competitive effects during the selection of bound morphemes increases as the paradigm (the PWI paradigm seems to be more insensitive in behavioural terms to detect the effect than simple naming tasks) becomes simpler, and as the size of the response set (most likely in regards of the number stems) becomes greater. Therefore, the ambiguity of the GCE for bound morphemes may be related to task-dependent reasons and to the inherent incapability of the PWI paradigm to detect certain effects, since it involves the processing of different elements through the comprehension and the production systems (and consequently, smaller competitive effects [which may be the case for bound morphemes in comparison to freestanding forms] cannot be easily picked up in this

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

paradigm).[13] Therefore, we believe that concluding that bound morphemes are not subject to selection by competition due to the absence of evidence obtained in three specific studies (Costa, Kovacic, Fedorenko, et al., 2003; Schiller & Caramazza, 2003; Schiller & Costa, 2006) is far from desirable.

5.1.2 Studies with Romance languages

Studies conducted with native speakers of Romance languages challenge what seems clear for Germanic and Slavic languages. On the one hand, when participants name the pictures by using noun phrases, there is an absence of any kind of gender effect in Catalan, Italian and Spanish (Catalan: Costa et al., 1999; Italian: Cubelli et al., 2005; Miozzo & Caramazza, 1999; Miozzo et al., 2002; Spanish: Costa et al., 1999; Miozzo et al., 2002). On the other hand, results are mixed when it comes to the production of bare nouns (Cubelli et al., 2005; Finocchiaro et al., 2011; O'Rourke, 2007; Paolieri, Lotto, et al., 2010; Paolieri et al., 2011).

The absence of effects with noun phrases has been addressed by the Late Selection Hypothesis (Miozzo & Caramazza, 1999), which describes this issue as a matter of mapping between each gender value and the specific number of determiner forms associated with each value. More precisely, there are languages that have more than one definite article per gender value, as is the case in Italian, which features the feminine determiners "*la*" and "*l'*" and the masculine determiners "*il*", "*lo*", or "*l'*". Their selection depends on the initial letter of the next word, which can be either a noun or a prenominal adjective (e.g., "*il divano*" [M, the sofa] vs. "*lo zaino*" [M, the backpack] vs. "*il grande zaino*" [M, the big backpack] vs. "*l'orologio*" [M, the watch]; something similar occurs in Catalan, French and Spanish, see Costa et al., 1999). Conversely, other languages have a one-to-one correspondence, as is the case in Germanic languages (e.g., in Dutch "*de*" [C] and "*het*" [N]). In languages with one-to-various gender-form mappings (e.g., Italian), the form of the determiner cannot be encoded until all the major constituents of the phrase (not only nouns but also possible adjectives) are phonologically specified and ordered within the sentence to be produced. Therefore, after gender access, the selection of a determiner form needs to 'wait' until this information has been processed and made available, which masks any possible competition arising due to the activation of the

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

[13] Note that Schiller and Caramazza (2003) used two adjectives (two stems) and still failed to obtain the GCE with bound morphemes. Importantly, as noted in the main text, the competitive effects generated by bound morphemes may be smaller than those by freestanding forms. Considering the apparently low sensitivity of the picture-word interference paradigm to capture certain phenomena of lexical access, the null results in this study may be due to the insufficient capacity of the task to guarantee the consistent observation of small effects.

distractor's gender (and the related determiner forms). Hence, in those languages in which the determiner is selected late, the GCE is not visible, whereas in languages with one-to-one gender-form mapping, an "early selection" of the determiner is possible and hence the effect is observable. Because there seems to be a clear distinction between the Romance and the Germanic/Slavic branches in this regard, Costa and colleagues (1999) defined the former as late-selection and the latter as early-selection languages. However, as pointed out by Foucart and colleagues (2010), this may be an overstatement that lacks enough precision to capture the complexity of the phenomenon. We agree, because: (1) there is at least one Romance language, Portuguese, in which the mapping between gender and determiners is one-to-one, and hence should behave like an early-selection language; (2) we are unsure about the fact that a language can be uniformly classified as early or late selecting. The claim for the uniform classification of a language as early or late gained popularity due to a study by Costa et al. (1999), who conducted several PWI tasks with Spanish participants and failed to obtain any effect of gender congruency. In Spanish, masculine nouns are always preceded by the definite article "el" (e.g., "el reloj" [the watch]) but feminine nouns can be preceded either by "la" (e.g., "la casa" [the house]) or "el" if the head noun starts with a stressed "a-" (e.g., "el alma", "el arma", "el águila" [the soul, the gun, the eagle]). Costa et al. (1999) failed to find any sign of a GCE in Spanish, even for the masculine. Importantly, the cases in which feminine nouns are preceded by "el" constitute only about the 0.5% of feminine occurrences (Caramazza et al., 2001). Consequently, they propose what is called the "maximum consistency principle" or "highest common denominator principle", contrary to what Miozzo and Caramazza (1999) defined as the "temporal optimization principle". According to the latter, the elements forming a noun phrase are organized for output as soon as possible. If this were true, the elements for a masculine noun phrase in Spanish should be organized as fast as in German or Dutch, since the article "el" does not depend on the form of the noun. This would mean that a GCE should have been observed for Spanish masculine nouns, which was not the case. Costa and colleagues thus argue that if in a language there are some determiners whose selection within the same gender value depends on the phonological form of the next element, all the determiners of that language will be selected at the same point in time, even if they do not depend on the phonological context,

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

preserving maximum consistency within lexical access. The null results in Spanish were replicated by Miozzo et al. (2002).[14]

There is one language that presents a more complex picture here, and that offers new insights into both the GCE and the PWI paradigm. Unsurprisingly, in light of the role it has played in many parts of this thesis, that language is French. French is a Romance language, in which the definite articles for each gender value are "le" (M) and "la" (F). Yet they both adopt the form "l'" when the next word starts with a vowel. Although Alario and Caramazza (2002) failed to find any effect of gender congruency in French with noun phrases formed by definite articles and nouns starting with a consonant, Foucart et al. (2010) obtained the effect in three different experiments when using a positive SOA of 200 ms. In their third experiment, the effect was obtained even when they included extra nouns starting with a vowel. Foucart and colleagues point out that in French the choice between the normal vs. the shortened determiner form does not refer to gender, since both masculine and feminine articles are delivered as "l'" before a vowel. Thus, the authors claim that it is not clear whether "l'" needs to be represented as a distinct determiner or whether, instead, it should be treated as a reduced form of the other determiners, resulting from a late phonetic adjustment. If the latter is true, the authors suggest that "le/la" would always be selected "by default", but full selection would occur after the form-based local context rule has been checked. This delay would slow down the selection of the determiner in French in comparison to Germanic languages, but it would still be faster than in certain Romance languages. With positive SOAs, the local context should have had enough time to be checked, and hence the competition between determiners activated by picture and distractor should result in variations in the naming latencies. The results by Foucart and colleagues show exactly this, and make relevant the idea that the observation of the GCE with noun phrases is a matter of gender-form mapping rather than of language families. Nevertheless, Bürki et al. (2020) recently failed to replicate Foucart et al.'s (2010) experiments, using the same and other materials. They also tested indefinite articles ("un" [M] vs. "une" [F]), which according to the maximum consistency principle should either not show any effect or show them in a positive SOA (just like the definite articles), but according to the temporal optimization principle, and the idea that it is, rather, a problem of mapping within a specific

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

[14] O'Rourke (2007) also failed to find effects. However, she used a SOA of 0 ms with an auditory presentation of the distractors, which we have seen may not be the best option to capture an effect, due to the delay between the gender selection of the target and that of the distractor. Hence, we cannot take O'Rourke's results as supporting evidence for the maximum consistency principle.

type of determiners, they should show them earlier (at SOA 0). However, the authors only tested a positive SOA of +200 ms, and the effect they managed to obtain was of only 9.6 ms (Experiment 1b, in contrast to the effect of +35 ms obtained with definite articles by Foucart et al., 2010). Besides, they applied Bayes factors and found that this was not a robust effect. A replication of the study showed null effects with both SOA 0 and +200. However, in their third and fourth experiments Bürki and colleagues decided to try a new method to look for the GCE with definite articles. Besides testing the SOAs 0 and + 200 ms, they considered that the PWI paradigm might be problematic due to the conjunction of picture naming and word reading. According to their reasoning, the SOA of the distractors could have a different impact on each participant, depending on how fast they could read, and that may explain why Foucart obtained the effect at +200 whereas they themselves did not. Ultimately, this could also open a new window on our understanding of the slipperiness of the small effects explored through this paradigm. Hence, they assessed the reading performance of their participants and considered it when analysing the naming latencies of the target nouns. More specifically, they measured the total reading time and number of errors of each distractor word of the list of stimuli. They computed the difference measure of processing time for distractor and target words by subtracting the distractor reading latency from the target picture naming latency for each picture-word combination. The measure was then centered around the mean. When, in the final model, they accounted for the variance in the difference measure along with the gender congruent and incongruent conditions, they obtained a GCE in both Experiments 3 and 4. The effect was independent of the SOA. Due to these results, they conducted a small meta-analysis with Foucart's data, Alario and Caramazza's (2002) data, plus data from another study by Schriefers and Teruel (1999), along with their own data. The results suggested positive evidence in favour of gender congruency with French definite articles, with no significant effect of SOA. They concluded that GCEs in French are not as robust as in other languages but may surface under certain conditions related to participants' processing times. The key is that GCEs in this language, and perhaps in other late-selection languages, are not simply restricted to late SOAs, but mainly depend on participants reading times. The Late Selection Hypothesis implicitly states that the distractor could still affect the RTs if it were still active at the time the target is accessed. Adjusting the reading times to the target and distractor processing times may allow us to observe effects in Catalan, Italian and Spanish. If researchers consider the variability in processing times for target and distractor words, the masking of the effect may be avoided, and

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

we may find a solution for the discrepancies found in many other studies with Romance languages. Critically, many of these discrepancies are actually related to the production of bare nouns.

Although the results for the production of noun phrases seem to be partially well-grounded in theoretical terms, when it comes to bare nouns the literature is difficult to explain. Findings are disparate and highly controversial. Whereas some studies have found null results of gender (Italian and French: Finocchiaro et al., 2011; Spanish: Finocchiaro et al., 2011)[15], others have found effects in an unexpected direction with native speakers of Italian and Spanish (i.e., a gender incongruency effect [GIE]: faster responses to targets paired with distractors of a different gender; Cubelli et al., 2005; Paolieri, Lotto, et al., 2010; Paolieri et al., 2011). Indeed, these effects obtained with bare nouns in Romance languages seem to be elusive, and the factors that affect their occurrence, although of great interest, are unclear. Importantly, neither WEAVER++ nor the IN model are able to accommodate this intricate pattern of data. Note that certain authors have interpreted the GIE as a genuine effect of gender and hence as support for the GSIH because, in opposition to the "GCE" obtained with Germanic and Slavic languages, the GIE is observed in the absence of elements of agreement (see Heim, Friederici, et al., 2009). However, in this thesis we will diverge from this view since we think that the origin of this effect is unclear. Above all, it seems highly unlikely that the selection of a gender node is hampered due to the existence of two sources of activation converging in this node (i.e., it is unlikely that a higher amount of activation would entail more difficulties to reach the threshold for selection). Besides, genuine effects of gender supported by competition cannot be explained as a matter of a gender node competing with itself. Thus, the GIE is more likely to somehow reflect competition between lemmas connected to the same gender node. Yet the mechanisms behind this competition are not clear. In any case, we believe that although the GIE does not reflect competition between gender nodes, this cannot be taken as supporting evidence for the non-competitive nature of gender either. Thus far, it is true that, with the exception of the fMRI studies (Heim, Eickhoff, et al., 2009; Heim, Friederici, et al., 2009) and a marginally significant result with bare nouns in Dutch (Experiment 2 of La Heij et al., 1998), competition between gender values or nouns due to gender seem to be either non-existent (DSIH) or extremely hard to capture, at least in Germanic and Slavic languages, and even in French. Yet both the WEAVER++ and IN models were originally (and still are) thought to accommodate

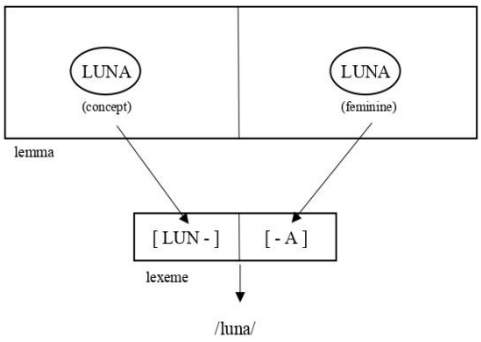
[15] We will not consider the study of O'Rourke (2007) here due to the auditory presentation of the distractors.

competition between gender nodes. If grammatical gender is indeed not subject to competition, we have seen that both models would have to modify their tenets. The GIE would also require other type of modifications to occur, but we will address them later in this subchapter. Right now, we are facing a situation that shows nouns competing for selection on the basis of gender and also of the language.

The fact that some languages show competitive effects between nouns based on gender and others do not, leads us to wonder what could be the diverging point between them. A peculiarity regarding Romance languages that is not considered by these two models stands out: gender transparency. As we have seen, Italian and Spanish are characterized by a high degree of gender transparency, with a majority of transparent nouns in their vocabularies. Cubelli and colleagues (2005) hence proposed the Double-Selection Model (DSM, see Figure 8) with the sole purpose of assessing in detail gender processing in language production when it came to Italian (and in later publications, to Spanish, Paolieri, Lotto, et al., 2010). In fact, the DSM is the only known model to have focused on the feature of grammatical gender, incorporating a simplistic notion of phonological gender transparency in its postulates. The authors based their theory on the morphological decomposition of nouns within the lexical organization proposed by Levelt and Schriefers (1987) and Levelt (1989), hence merging the first two levels of lexical access, the semantic and the syntactic strata, into a unique level represented in the lemma itself. Thus, in the DSM the syntactic information is not represented as a separate type of information mediating between semantics and phonology; rather, it is an intrinsic part of the lexical representation of nouns. Grammatical gender is therefore not represented as gender nodes, and at the lemma level, each noun is specified for both semantics and syntax, with their gender value as an extra feature. All the semantic and syntax information concerning one noun has to be selected for the morpho-phonological information to be accessed, and thus the authors postulate a double selection mechanism in which the

Figure 8

Lexical access according to the Double-Selection Model



Note. Production of the noun “luna” (moon) in Italian. Adapted from Cubelli et al. (2005).

semantic representation is selected prior to the syntactic information.

Cubelli and colleagues argue that this proposed architecture of lexical access is universal, but the underlying mechanisms are constrained by the peculiarities of each language. By relying on the notion of morphological decomposition, the authors claim that semantics and syntax are two parallel-related levels that connect to the form stratum, in such a way that different information at the lemma level activates different components of the word form separately. For Romance languages, semantic information allows for noun stem selection, and grammatical properties entail access to the correct nominal ending and the associated declensional class. Their reason for the compulsory nature of the selection of grammatical gender for word form to be accessed is the complex morphological structure of Italian nouns. In this language, grammatical gender and number coincide, since the morphological marker for plural depends on the grammatical gender value itself (e.g., the Italian noun "luna" [moon] is feminine, and takes the morpheme "-e" for plural). Hence, they consider that transparent cues are actually "gender marks" that depend on gender selection. For instance, the noun "luna" would encompass the processing of semantics for the selection of the root form "lun-" and the processing of the feminine gender value plus the singular for the processing of the form "-a". In the case of Italian (and Spanish) gender is selected and gender competitive effects are observable. Importantly, competition occurs by means of similarity, both semantically and syntactically. Hence, the semantic interference effect and the GIE would operate by the same mechanisms, in which nouns (targets and distractors) of the same semantic class and the same gender would activate each other and compete for selection. Conversely, for opaque languages, gender is not necessary to process the nominal ending of nouns, and hence gender is not selected, and gender competitive effects are absent.

Although this proposal is certainly the first to consider the transparency of nouns and the possible modulations on grammatical gender representation and processing depending on the characteristics of the language itself, it has led to a series of criticisms in the literature (Finocchiaro et al., 2011; Sá-Leite et al., 2019). One of the main concerns arises from the basic tenets in which the model is grounded. Due to the relation of dependency that exists between gender and form (and semantics and form), it seems that the authors dispose of the principle of independence between lemma and lexeme. This goes against most of the literature on lexical access. Regarding grammatical gender, there are many studies showing the relation of independency between gender and word form through the assessment of brain-damaged

USC UNIVERSIDADE DE SANTIAGO

patients. For instance, Badecker et al. (1995) reported the performance of an anomic subject, Dante, who despite his inability to produce the names of objects, or even guess the first or last phoneme in a two-phoneme, forced-choice task, was always able to recall their gender correctly. There is the possibility that if indeed the last phoneme of a word is directly computed through the gender value, the recovery of gender would entail the recovery of the last phoneme according to the relation of dependency apparently established by the DSM. Many other studies show that access to a word's phonological features can occur independently of access to its syntactic features. Henaff Gonon and colleagues (1989) were the first to observe that brain-damaged subjects with anomia seemed to know the gender of nouns that they could not produce. Miozzo and Caramazza (1997a) observed that a brain-damaged anomic subject could retrieve the gender of nouns but was no better than chance at choosing the phonological features of known words that he could not name. There are also reports of an Italian speaker, FS (Miceli & Caramazza, 1988), who produced utterances in which the form of nouns were correctly encoded but the preceding articles were incorrect in terms of gender. Miozzo and Caramazza (1997b) and Vigliocco et al. (1997) showed that the retrieval of gender information in Tip of the Tongue (TOT) states did not interact with the regularity of the nouns' endings. In fact, in some experiments participants seemed to be more accurate in retrieving the initial than the final letter after successfully recovering gender. In a more precisely focused study, Caramazza and Miozzo (1997) evaluated the relation between positive TOTs (the ones in which the word that the participant had in TOT state is the correct one) and the degree of accuracy in retrieving partial phonological information of the word in TOT. The words were either transparent or non-transparent (opaque or irregular). Similar to previous studies, they found that the regularity of a word's ending had no effect on the selection of gender in the TOT condition.

Another concern regarding the DSM model comes from the morphological decomposition that Cubelli et al. (2005) and later Paolieri, Lotto, et al. (2010) and Paolieri et al. (2011) claim speakers of Romance languages perform. Particularly, phonological cues are considered to be gender morphemes in all these studies. However, this seems quite imprecise. According to the authors, a Spanish word such as "casa" [house] would be decomposed in "cas-" and "-a", with "-a" standing as a gender marker whose retrieval depends on gender processing. This would be similar to what would occur with the word "chica" (girl), which does not have grammatical gender but, rather, natural gender. In "chica" the feminine value is marked through the gender morpheme "-a" in opposition to "-o" in the word "chico" (boy). However, in "casa", "-a" is a

USC UNIVERSIDADE DE SANTIAGO

thematic vowel, such as "-o" in "codo" (M, elbow) and "-e" in "torre" (F, tower). These are not morphological gender markers. Instead, they emerge from the evolution of the Latin declinations of the words from which today's nouns proceed (Vilela, 1973). Nowadays, nouns ending in "-a" tend to be feminine and the ones ending in "-o", masculine. However, there are exceptions, and so the authors (Cubelli et al. 2005, Paolieri, Lotto et al., 2010; Paolieri et al., 2011) leave the processing of irregular nouns unexplained (in which the encoding of one gender value would entail the computation of the wrong nominal ending). Similarly, there are also many opaque nouns that are athematic ("luz" [light]) and that have the thematic vowel "-e" ("torre" [tower]), and the authors do not explain how word form is retrieved in these cases. However, if gender is retrieved for all the nominal endings to be selected except when they are not transparent, one would assume either that: (1) speakers would inevitably make many errors of word-form encoding during normal speech (which has not been reported in the literature); (2) gender retrieval for opaque and irregular nouns would function as gender retrieval for opaque languages. In any case, in relation to the two points above, we might observe the following: (1) it is unclear how the system is able to identify which nouns depend on gender retrieval for word-form encoding and which nouns do not, if these transparency cues for gender are not available (and if they depend on the previous retrieval of gender); (2) the GIE has been observed for both transparent and opaque nouns (Cubelli et al., 2005; Paolieri, Lotto, et al., 2010; Paolieri et al., 2011).

Despite the inconsistencies, the authors discussed above insist that their rationale is accurate for Italian because its complex morphological system intertwines gender with number: number morphemes in the plural depend on the grammatical gender value itself. For instance, and similar to Spanish, house in Italian is "casa" (F), but whereas in Spanish it takes the morpheme "-s" for plural, as do other feminine and also masculine words, in Italian it takes the morpheme "-e" for plural, "case". Masculine nouns, on the contrary, take the morpheme "-i" for plural (e.g., bed in singular is "letto", but in plural is "letti"). Due to the existence of different plural morphemes that depend on gender ("-e" [F] vs. "-i" [M]), Cubelli and colleagues state that the grammatical encoding of gender is always necessary for speakers to know the ending of a noun, even when (1) the noun is opaque or irregular, because the plural is transparent, and (2) participants do not intend to produce words in the plural. Nonetheless, this idea is debatable. There is solid evidence in favour of the singular-as-default hypothesis for Dutch, German and French (Alario & Caramazza, 2002; Janssen & Caramazza, 2003; Lemhöfer et al., 2006;

USC UNIVERSIDADE DE SANTIAGO

Schriefers et al., 2002; 2005), which shows that our brain works in the singular unless the plural is overtly necessary in terms of semantics (e.g., if cued by the researcher in a naming task – when the plural is produced, the plural forms compete with the singular forms of determiners, but the opposite is not true). If we do indeed function cognitively in the singular by default, and the endings of singular nouns do not depend on a gender value, it is unclear why Italian speakers would process gender in the singular to obtain a nominal ending. Besides, the authors extend this model to all those languages that are highly transparent. However, Portuguese and Spanish do not adhere to this complex system of morphological rules that link number and gender. Despite this, gender effects have indeed been obtained in another Romance and highly transparent language such as Spanish (Paolieri, Lotto, et al., 2010).

5.1.3 Discussion on native language production

To draw conclusions on the GCE, we will consider two sets of studies: those conducted with native speakers of Germanic and Slavic languages, and those with native speakers of Romance languages. In the former, we mostly find that competition between gender congruent and incongruent conditions only occurs when participants produce noun phrases formed by freestanding determiners whose form varies across gender values (i.e., the "GCE"). In the latter we mostly find an absence of competitive effects with noun phrases and either no effects or effects in the opposite direction when it comes to bare nouns (i.e., the GIE). However, behind this apparently "simple" description of results, there is a great deal of mixed evidence that must be accounted for in order to make a rigorous assessment of the mechanisms underlying gender selection and the GSIH vs. DSIH discussion.

Regarding Germanic and Slavic languages, evidence mainly supports the DSIH since there is a systematic absence of a GCE when determiners do not vary across gender values, i.e., ordinals with soft declension in Czech, plural definite articles in German and Dutch, and articles for diminutive noun forms in Dutch (note that in the case of diminutive noun phrases, in line with the DSIH, the effect is obtained in the inverse direction; Bordag & Pechmann, 2008; Schiller & Caramazza, 2003, 2006). Although we could argue that because an agreement context is absent, gender is not being selected, and hence competition for selection does not occur, studies on the singular-plural paradigm show that this is not the case at least when it comes to the plural and diminutive forms. More specifically, there is a cost for words whose gender have different determiners in the singular and the plural, as well as in the diminutive

USC UNIVERSIDADE DE SANTIAGO

and the standard forms, but a facilitation when the determiner is the same. These results suggest that, indeed, gender is being accessed during the production of plural and diminutive noun phrases even if the determiner has the same form for every gender (i.e., if agreement is not being fulfilled). Hence, if the GCE is not observed, it seems to be due to the absence of different forms of determiners for targets and distractors.[16] Another point in favour of the DSIH concerns the null results obtained with bare nouns, which come from the studies of La Heij et al. (1998), Starreveld and La Heij (2004), and Finocchiario et al. (2011; but see the second experiment of La Heij and colleagues [1998], in which they obtained a marginal GCE which was only significant by subjects with Dutch participants). Here again, agreement forms are not being produced and the effect is absent.

Certainly, this body of evidence seems sufficient to locate the so-called GCE at the level of the phonological encoding of determiners, hence accepting the DSIH when considering Germanic and Slavic languages. However, this is all we are able to say with the evidence we currently have. The other claims that have been linked to the DSIH are highly debatable. More specifically, we find it quite problematic to argue that, because what can be measured in the PWI paradigm is a determiner effect, then gender is an abstract feature that is automatically retrieved and not subject to competitive effects, which would make gender experimentally not observable (Caramaza et al., 2001).

In relation to this claim, there is enough data in the literature at least to soften it. When inspecting the RTs of Schiller and Caramazza (2003) with singular forms in comparison to plural forms, RTs are overall lower for the latter. The same happens with Schiller and Caramazza's studies (2006), in which the diminutive forms produce clearly faster responses. There are two important issues that should be noted regarding these results:

- As has been consistently shown in many experiments with the singular-plural paradigm, plural and diminutive forms must show lower RTs as a function of gender (e.g., Janssen and Caramazza, 2003; Schriefers et al., 2002; Schriefers et al., 2005; Spalek & Schriefers, 2005). Only the gender value using the same articles or inflections in both forms (e.g., singular and plural) should show these lower RTs, and this should not occur for the other gender values, in which a cost should be observed. Consequently, the PWI

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

[16] Schiller and Caramazza (2003) show that when the target has to be produced in its diminutive form, the distractor activates the standard form of the word, but when the plural is required, both the target and the distractor are interpreted in the plural form. The number of competing nouns does not seem to be considered and it seems a feature that is determined extra-lexically according to the intentions of the speaker.

paradigm does not seem to properly reflect this competition that occurs between determiner forms of the same gender during the selection of a noun in plural or in the diminutive form.

- The lower RTs in the plural and diminutive conditions may mask differences related to gender selection (Heim, Friederici, et al., 2009, p. 937).

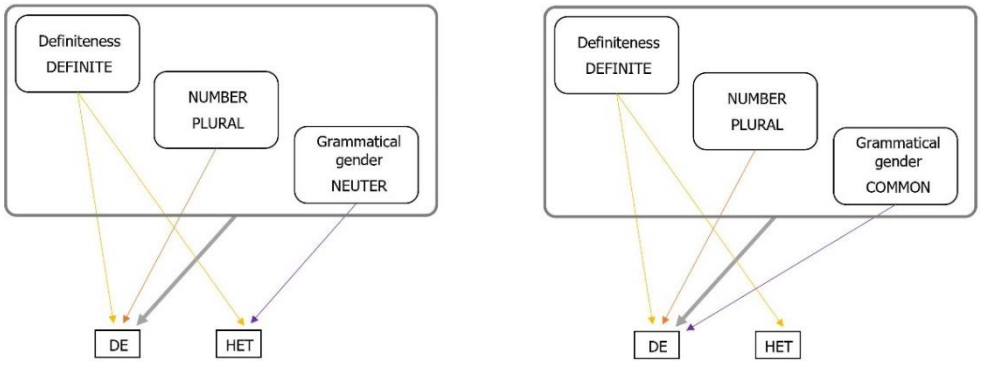
It would seem as if the PWI paradigm is not adequate to examine any effects produced by competition between lexical and syntactic features. The studies on the singular-plural paradigm, along with others such as Alario and Caramazza (2002) study, seem to be a perfect reflection of this. These studies form a strong body of evidence on describing the way determiners are selected. More specifically, they support the *Primed Unitized Activation Hypothesis* (see Janssen & Caramazza, 2003) in clear detriment to the Hierarchical Selection Hypothesis and the Unitized Activation Hypothesis (see Alario & Caramazza, 2002). The Primed Unitized Activation Hypothesis (see Figure 9) states that the features responsible for the selection of a certain determiner spread cascaded activation. For instance, during the production of the Dutch plural noun phrase "de huizen" (the houses, from the neuter noun "het huis"), each of the features, plural and neuter, send activation to their associated determiner ("de" and "het", respectively). However, the unitized representation of both features, i.e., "plural + neuter", also sends activation to its associated determiner ("de"). In the case of the Dutch noun phrase "de tafels" (the tables, from the common noun "de tafel"), each feature would activate "de", and the unitized representation of both features would also activate "de". This hypothesis hence predicts that the gender of the noun should contribute to certain variation in the naming latencies of plural noun phrases (i.e., a cost for neuter gender; a facilitation for common gender). The same would be predicted for diminutive noun phrases. Since the effects of the Primed Unitized Activation Hypothesis are clearly reflected in the singular-plural paradigm, it is quite interesting how they seem to be masked in the PWI paradigm, in which activation would be cascading not only from the features of the target noun, but also from the features of the distractor noun in either the plural or the singular. Yet neither the classical GCE nor the null results obtained by Schiller and Caramazza (2003) in the plural wholly reflect this. Importantly, as the activation cascades from one feature to another (definiteness, number, and gender), the last competitive process occurs between articles and involves activation deriving from multiple sources. However, gender competition takes place before that and involves only two sources of information restricted to the feature of gender itself (e.g., common vs. neuter). This being true,

USC UNIVERSIDADE DE SANTIAGO

it is not surprising that gender competition is masked by other processes involving higher levels of competition and the only way to observe gender effects would hence be through bare nouns. Yet if gender competition entails small effects due to a low degree of activation coming from gender nodes, it does not come as unexpected that the PWI paradigm is not able to capture such effects.

Figure 9

Primed Unitized Activation Hypothesis



Note. Representation of the plural definite determiner selection in Dutch for neuter nouns (left) and common nouns (right). Adapted from Alario and Caramazza (2002).

Supporting these concerns is the literature on another issue on which researchers do not seem to have reached a consensus: that of the mechanisms underlying the selection of freestanding elements and bound morphemes. The final claim accompanying the DSIH is that the determiner congruency effect occurs exclusively with freestanding forms because bound morphemes are not selected through competitive mechanisms. However, once more it seems that the PWI paradigm is not successful enough at reflecting competitive effects between bound morphemes of different gender, as shown mainly by Jescheniak et al. (2014). All in all, we suspect that this may be due to the clearly shorter RTs in studies with bound morphemes compared to those with freestanding morphemes. More specifically, when it comes to bound morphemes, authors tend to include only one possible element of agreement (i.e., the same stem) and various suffixes, which seems to be speeding up the moment at which the participant begins to respond by enabling the use of the same first word in their response. This tendency is reduced when more than one element of agreement (more than one stem, for instance, two stems belonging to two adjectives) is introduced as a possible naming answer. When this tendency decreases, the "GCE" is obtained. Evidence on the singular-plural naming paradigm confirms these results, in which effects of competition with bound morphemes are obtained in the same

manner as freestanding determiners, except when only one stem is used in the naming response (Jescheniak et al., 2014). Hence, shorter RTs also decrease here the probability of the PWI task reflecting competitive effects.

In sum, if competition between phonological forms, which seems to be fairly easy to capture, can be masked within the processing involved in the PWI paradigm (singular vs. plural forms; bound morphemes), small competitive effects regarding syntactic features (i.e., gender) may be even harder to observe. Hence our concerns as to the conclusions that can be drawn regarding the nature of gender: does the absence of gender effects on this paradigm truly constitute proper evidence to claim that gender values do not compete?

In relation to the probable insensitivity of the PWI paradigm to reflect certain small effects, we may ask ourselves what would happen when we use more fine-grained techniques. As we have seen, Heim, Eickhoff, et al. (2009) and Heim, Friederici, et al. (2009) used fMRI measures and found supporting evidence for the activation of gender, in that gender incongruent conditions showed slower activation than gender congruent conditions in the posterior inferior part of the left BA 44 area (Broca), both with noun phrases in the PWI paradigm, and with bare nouns in a naming task with priming. This area has previously been associated with the processing of abstract syntactic features, rather than with morpho-phonological encoding and the retrieval of determiners. The study by Bürki et al. (2016) however, seems to be less clear in its outcomes. Their results with EEG measuring techniques has been interpreted as evidence for the DSIH, but we are unsure as to what extent their data can truly discriminate between the DSIH and GSIH hypotheses. The authors state that if the "gender" congruency effect were reflecting gender processing, it should be shown in earlier time windows than a phonological facilitation effect. Yet the inverse outcome was obtained. Therefore, the authors state that the "gender" congruency effect in fact reflects the selection of determiners, and that the form of the nouns is thus encoded before the determiner forms, which they argue are accessed late during lexical access. In any case, they cannot adequately explain an even later effect of phonological overlap taking place after the "gender" congruency effect. In theoretical terms, the interpretation of this late GCE as evidence for the DSIH is in line with WEAVER++, in which the lemma encoding precedes that of the lexeme. However, if we adopt the view of the IN model, gender processing would be occurring mainly after the encoding of the noun lexeme, perhaps with a certain degree of overlap (the IN model argues for the existence of a cascading flow of activation, suggesting the possibility of some parallel activation or overlap). This

USC UNIVERSIDADE DE SANTIAGO

invalidates the interpretation of the late GCE as evidence in favour of the DSIH. The same would be the case with a more interactive and cascaded view of the WEAVER++ model, in which the lexeme and the syntactic properties of a noun may be processed in parallel. To correctly discriminate between a late competitive effect of gender vs. of determiner form, the authors would need to include a condition with bare nouns as a naming instruction.

Interestingly, considering the literature on Romance languages, this late "gender" congruency effect obtained by Bürki seems to suggest that indeed, either gender, determiner forms, or perhaps both, are processed later than the phonological encoding of the noun. This view may seem incompatible with the tenets of the Late Selection Hypothesis, according to which the phonological encoding of the noun occurring previously to the selection of the correct article would be the main source of the masking of any competitive effect. However, if Bürki and colleagues (2016) results are indeed replicable, the Late Selection Hypothesis can still maintain its rationale if we adjust exactly what is being encoded before article selection can take place. In Italian, extra information about the context of the determiner is also necessary, because the inclusion of an adjective between the determiner and the noun may change the article required by the noun ("lo sgabello" [the stool], but "il grande sgabello" [the big stool]). In Spanish, besides taking into consideration possible adjectives ("el águila [F]" [the eagle], but "la majestuosa águila" [the majestic eagle]), the speaker also has to process the stressing patterns of nouns, since only those starting with a stressed "-a" require the article "el" ("el **a**gua" vs. "la a**cel**ga" [the water vs. the chard]). Thus, in these languages, besides the phonological encoding of the noun, additional processes of word organization and of syllabification also have to take place before a determiner can be selected. Taking into account the results of Bürki and colleagues (2016), these additional processes and not the phonological processing of the noun form would be responsible for the masking of the GCE.

Indeed, the general absence of effects with noun phrases in speakers of Romance languages makes it even harder for us to obtain any evidence either for the GSIH or for the DSIH. An interesting language is French, because despite the fact that the majority of studies fail to obtain the GCE (Alario & Caramazza, 2002), there is at least one, conducted by Foucart et al. (2010), which obtained the "gender" congruency effect in French with positive SOAs. Although Bürki et al. (2020) repeatedly failed to replicate these results, they found something quite interesting. The authors suggest that the determining factor in observing the GCE might be the matching of the reading times of each participant with their respective naming times. Indeed, these authors

USC UNIVERSIDADE DE SANTIAGO

do find significant GCEs in French independently of the SOA, and claim that GCEs may be hidden by a notable variability in processing time differences between picture naming and distractor processing.

Bürki et al. (2020) also point out something that we find to be quite important: if there is so much variability in participants' reading times, it is possible that with a far larger cohort of participants, the interaction between gender conditions would have a greater possibility of reaching significance. Although this is a statistical matter, it is in fact very pertinent to the literature in this area. A brief review of the studies discussed in this chapter shows that most of them do not exceed 30 participants per experiment, not even Bürki and colleagues' (2020) study, in which the largest sample they used was of 30 participants, in their fourth experiment (in which the GCE was obtained; in their experiments, however, they use a greater number of stimuli than in most of the behavioural literature in the area: 48 pictures). In Germanic languages, only the study by van Berkum (1997; Experiment 2) and Bordag and Pechmann's first experiment (2008) used more than 30 participants (n = 48, n = 32, respectively). For Romance languages only Paolieri et al. (2011) do so (n = 36). Strikingly, many studies use fewer than 20 participants using similar experimental designs (Bordag & Pechmann, 2008 [Experiments 2 and 3]; Finocchiaro & Caramazza, 2006 [Experiments 2, 3A, and 3B]; Foucart et al., 2010; Miozzo & Caramazza, 1999; Miozzo et al., 2002; O'Rourke, 2007; Paolieri, Lotto, et al., 2010 [Experiment 1]; Schiller & Caramazza, 2003 [Experiments 2a, 2b, 4a, and 4b], 2006 [Experiment 2]; Schriefers, 1993; Schriefers & Teruel, 2000). Of great concern here are the studies by Bürki et al. (2016) with only 18 participants (although a higher number of pictures, 70), and by Heim, Friederici, et al. (2009) with only 16 (60 pictures). These studies use more fine-grained techniques but would have been more reliable with larger samples of participants, especially when considering the importance of their outcomes in the field. Indeed, a small sample size is worrying, but is even more so if we take into account the very probable low sensitivity of the PWI paradigm to reflect certain competitive effects, especially if these are small and prone to be affected by multiple external or internal factors (e.g., such as other processes taking place during lexical access [Late Selection Hypothesis], variability in reading times, etc.). In this sense, another factor that could influence the results on such a paradigm is the control of the stimuli used in the experiments. A quick inspection of the reviewed studies reveals that certain psycholinguistics parameters have never been considered in this field despite their relevance for picture naming or word (distractor) recognition, namely the number

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

of phonological and orthographic neighbours (Grainger et al., 1989; Zhang et al., 2020) and the imageability (Perret & Bonin, 2019), concreteness (Soares, Lages, et al., 2019), and familiarity or subjective frequency (Qingfang & Yufang, 2003; Soares, Lages, et al., 2019), among others. We will address this in more detail when presenting our own studies in Chapter 6.

Finally, the coup de grace of this whole discussion concerns the GIE that has been reported several times in Spanish and Italian when participants produce bare nouns (Cubelli et al., 2005; Paolieri, Lotto, et al., 2010; Paolieri et al., 2011). On the one hand, this effect suggests that rather than gender nodes competing, nouns compete on the basis of gender. Yet this effect is not appropriate to assess the DSIH, since it is obtained with bare nouns – any elements of agreement that could engage in form selection competition are absent. In any case, if we want to maintain the widely accepted representation of grammatical features as nodes (i.e., gender as gender nodes) in a grammatical stratum along with other features (WEAVER++ or IN models), but also to predict an incongruent effect of competition between nouns depending on gender, some basic modifications are in order concerning the way gender nodes and nouns lemmas are conceived. We can either (a) include some kind of backward connection between the gender nodes and the rest of the noun lemmas of the lexicon, or (b) create a cohort of connected noun lemmas of the same gender competing for selection. The available evidence, however, is mixed, and although it has been suggested that bidirectional connections between gender nodes and noun lemmas exist (see Vigliocco et al., 2004), other studies only support bidirectional connections from gender nodes to other word lemmas in the case of other elements of agreement, in which gender marking is important/necessary (determiners, adjectives, numerals, etc.; see for example Jescheniak, 1999). In these cases, elements of agreement presented as primes (including noun phrases consisting of definite articles and nouns) have been shown to facilitate the processing of target nouns of the same gender in many languages, see e.g., Bates et al., 1996; Bender et al., 2011; Bölte & Connine, 2004; Deutsch & Wijnen, 1985; Guillelmon & Grosjean, 2001; Jacobsen, 1999; van Berkum, 1996 (doctoral dissertation). When bare nouns are presented as primes (instead of full noun phrases or elements of agreement), priming effects are either absent (e.g., Jescheniak & Schriefers, 2001; van Berkum, 1996) or are positive (facilitation for nouns of the same gender, e.g., Heim, Eickhoff, et al., 2009; Vigliocco et al., 2002). Importantly, as seen in the previous section, the DSM by Cubelli and colleagues (2005) does not seem to offer a robust theoretical basis that is consistent with the literature to explain either the appearance of effects due to transparency or the direction of these effects.

USC UNIVERSIDADE DE SANTIAGO

Importantly, we must remark that the GIE cannot be taken as evidence to either support or dismiss the GSIH. The GIE tells us that nouns themselves are competing for selection, whether or not gender nodes are competing in an earlier phase cannot be assessed. Indeed, it could be that the activation reaching a gender node from two different lexical entries transfers to the lemmas of the same gender, which end up involving higher levels of activation and thus being reflected as a GIE.

Currently, the answers to many of these issues are either non-existent or unsatisfactory. We will now turn to two different sources of information as a means of shedding more light on the debates surrounding the mechanisms underlying gender selection and the GSIH and DSIH hypotheses. We will first present a meta-analysis with the available data, towards finding some common ground on all the controversies presented here, and especially to check the size of the effects we are assessing, since their size when tested in a seemingly insensitive paradigm such as the PWI task is critical. Then, we will consider the results in the cross-linguistic GCE, which may provide us with other cues as to the nature of grammatical gender in terms of (1) its organization within the lexical access architecture, and (b) the mechanisms underlying its selection (competitive-based vs. automatic).

5.1.4 Meta-analysis of the monolingual data[17]

The "GCE" has been described exclusively as a determiner congruency effect for Germanic and Slavic languages. However, studies show inconclusive results regarding the type of agreement involved in the emergence of the effect. Conversely, in certain studies exploring Romance languages, the nouns of lemmas seem to be competing on the basis of gender. Yet null effects have also been obtained. Importantly, the complexity of the PWI paradigm has been pointed out as a possible obstacle to adequately detect small effects of competition, especially when it comes to gender. For this reason, authors have relied on a simpler task, the singular-plural paradigm, to draw conclusions in topics where the PWI paradigm has been incapable of doing so. These concern mainly a) the selection of gender during the processing of plural and diminutive noun forms, and b) the selection of bound morphemes. In the following pages, we will address quantitatively all these matters to determine the robustness of the effects of determiner forms and the lemmas of nouns in Germanic/Slavic and Romance languages.

USC UNIVERSIDADE DE SANTIAGO

[17] Based on the paper: Ana Rita Sá Leite, Karlos Luna (*Universidad Nacional de Colombia*), Ângela Tomaz (*Université de Lille*), Isabel Fraga, & Montserrat Comesaña. (2021*). The underlying mechanisms of grammatical gender selection in language production: A meta-analysis of the gender congruency effect* [Manuscript submitted for publication].

We will hence conduct a meta-analysis on the robustness of the "GCE" including all the studies that test Germanic and Slavic languages, as well as all those studies testing Romance languages but limited to bare nouns. We will then determine the size of the "GCE" in Germanic and Slavic languages, to confirm the DSIH and to assess its robustness depending on the presence of agreement (noun phrases vs. bare nouns) and the type of agreement element included in the naming phrase (freestanding vs. bounded). Regarding Romance languages, we will assess the size of the competitive effects when considering bare nouns. Due to the controversial results obtained for Romance languages, plus the number of studies in each language which have a low number of participants, we will also seek to assess a possible publication bias, this mainly to check whether there has been an overestimation of the GIE, but also to check the reliability of the determiner congruency effect in Germanic/Slavic languages. We then will conduct a smaller meta-analysis on the singular-plural paradigm. We intend to analyse the size of the competitive effects between elements of agreement of different type, gender, and number, and establish a comparison with the size of the GCE achieved by the PWI paradigm.

5.1.4.1 Method

5.1.4.1.1 Literature search

We conducted a comprehensive and systematic search of two online databases: APA PsycINFO and Web of Science (WoS). The search included the key words: "gender congruency", "grammatical gender", and "picture word interference", the first of these on its own, and the other two combined. We obtained 76 results from APA PsycINFO and 85 results from WoS.[18] From a total of 161 studies, we deleted duplicates with the RefWorks® citation software and obtained 100 different studies of interest. By reviewing the titles, the abstracts, the keywords, plus the full text of the studies where necessary, we applied the following criteria for inclusion:

a) The study constitutes an empirical exploration of grammatical gender representation and processing during the lexical access of nouns, and entirely avoids the inclusion of nouns with natural or semantic gender in the stimuli list (e.g., Deutsch & Dank, 2018).

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

[18] We excluded the following categories from the WoS database search: *Business, management, audiology speech language pathology, education research, psychology applied, psychology social, social sciences interdisciplinary, film radio television, health care science services, medicine general internal, otorhinolaryngology, public environmental occupational health, radiology nuclear medicine medical imaging*, and *urban studies*.

b) The study features one or various PWI paradigm tasks without variations in the classical procedure (i.e., fixation cross, optional blank screen, stimuli presentation). Hence, excluded are studies that used extra stimuli as primes (e.g., Alario et al., 2004) or linguistic instructions preceding the target stimuli (e.g., using a certain word [verb] or another depending on an indication written before the picture, see Finocchiaro & Caramazza, 2006).
c) Gender congruency was included as one of the experimental conditions.
d) The GCE explored in the study was restricted to the native language or dominant language of the speakers, and the study did not address cross-linguistic GCE, i.e., the effect that reflects the cross-linguistic interaction between translations of the same or opposite gender value during lexical access in bilinguals.
e) Participants were neurologically typical and did not present any language impairment (e.g., aphasias).
f) Participants were speakers of Germanic, Romance, or Slavic languages, since Semitic languages such as Hebrew (e.g., Dank & Deutsch, 2015) display a very different set of word-formation rules that may affect the comparability of results with the other language families.
g) Participants were asked to name pictures using either a bare noun, a noun phrase, or pronouns (this excludes any task variations in which participants had to carry out extra cognitive operations such as producing articles without nouns, e.g., Starreveld & La Heij, 2004).
h) Since the focus of our meta-analysis did not include exploring the Late Selection Hypothesis (Miozzo & Caramazza, 1999), the studies featured speakers of Germanic and Slavic languages with the naming instructions defined in point g, or speakers of Romance languages who used bare nouns to name the pictures (i.e., this excludes studies featuring speakers of Romance languages using noun phrases, e.g., Alario & Caramazza, 2002; Miozzo & Caramazza, 1999).
i) The aim of the study was not to explore grammatical gender selection in derived or compound nouns (e.g., Lorenz & Zwitserlood, 2016).

USC UNIVERSIDADE DE SANTIAGO

After the application of the inclusion criteria, we identified 12 studies of interest (Bordag & Pechmann, 2008, Bürki et al., 2016; Costa, Kovacic, Fedorenko, et al., 2003; Cubelli et al., 2005; La Heij et al., 1998; Paolieri, Lotto, et al., 2010, 2011; Schiller & Caramazza, 2003;

Schiller & Costa, 2006; Schriefers & Teruel, 2000; Scriefers, 1993; van Berkum, 1997). An inspection of the reference lists of these papers was subsequently conducted and other four studies were deemed eligible (Finocchiaro et al., 2011; Heim, Friederici, et al., 2009; O'Rourke, 2007; Schiller & Caramazza, 2006). Importantly, when several reports of the same study were encountered (e.g., journal article and dissertation: O'Rourke, 2007, 2009), we considered only the journal article version. All the authors in the field were contacted by email to request the raw data from their published works as well as unpublished data. Raw data was obtained from Cubelli et al. (2005), the fourth experiment of Finocchiaro et al. (2011), Paolieri, Lotto, et al. (2010), Paolieri et al. (2011), and van Berkum (1997)[19]. In total, 16 entries (all published papers) were carefully analysed, as described in Table 5.

To further address the debates on (1) the sensitivity of the PWI paradigm on the detection of competitive effects, and (2) the mechanisms behind the selection of bound morphemes, we conducted an additional, smaller meta-analysis of the singular-plural paradigm using only the search engine of APA PsycINFO. The search included the key word "bound morphemes", and yielded 58 results, and the RefWorks® citation software did not find any duplicates. We reviewed the papers and applied the following criteria for inclusion: (a) to preserve a direct comparison with the competitive effects obtained in the PWI paradigm, the studies explored Germanic or Slavic languages; (b) the studies used the singular-plural naming paradigm and bound morphemes or freestanding elements as naming phrase type; (c) since the effects cannot be taken together and understood as one, only manipulations on number were considered, not the manipulation on the standard vs. diminutive Dutch forms; (d) the studies were from the field of cognitive psychology and not developmental psychology, and thus the aim was not to study the acquisition of bound morphemes; (e) participants were healthy and not language impaired adults. Two studies matching the criteria were found: Lemhöfer et al. (2006) and Schriefers et al. (2005). An inspection of these led us to identify a further three important studies: Schriefers et al. (2002), Janssen and Caramazza (2003), and Spalek and Schriefers (2005). After contacting the authors, we were informed of another study of interest, Jescheniak et al. (2014). A total of 6 studies were therefore included in the analysis, and are described in Table 6[20]. Importantly, since the same effect must be measured across all papers, we focused on the facilitation effect found in the feminine (German) and in the common (Dutch) gender values.

USC UNIVERSIDADE DE SANTIAGO

[19] We thank Francois-Xavier Alario, Daniela Paolieri, and Jos van Berkum for sharing their data with us.
[20] We thank Kristin Lemhöfer for sharing her data with us.

Table 5

Summary of all reviewed studies on the GCE for the main meta-analysis

	Experiment	Participants (n)	Language	Language Family	Agreement context	Type of Phrases	SOA	GCE
Schriefers (1993)	E1	18	Dutch	Germanic	Y	NP (df + adj + N)	-200, 0, +450	Y -200, 0;
	E2	18				NP (adj + N)		Y 0
van Berkum (1997)	E2	48	Dutch	Germanic	Y	NP (df + N)	0	Y
La Heij et al., (1998)	E1	20	Dutch	Germanic	N/Y	BN/NP (df + N)	0	N/Y
	E2	20				BN/NP (df + N)		Y*/Y
	E3b	16				BN/NP (df + N)		N/Y*
Schriefers & Teruel (2000)	E1	16	German	Germanic	Y	NP (df + adj + N)	-150, 0, +150, +300	Y +150
	E2	16				NP (df + N)	-225, -150, -75, 0, +75, +150, +225, +300	Y +75
Costa, Kovacic, Fedorenko, et al. (2003)	E1	20	Croatian	Slavic	Y	Pron	0	Y
	E2	20				NP (adj + N)	0	N
	E3	19				NP (adj + N)	+75, +150	N
Schiller & Caramazza (2003)	E1a	27	German	Germanic	Y/N	NP (df + N) – S/Pl	0	Y/N
	E1b	25	German		Y/N	NP (adj + N) – S/Pl	0	N/N
	E1c	26	German		Y/N	NP (df + N + adj) – S/Pl	0	Y/N
	E2a	17	Dutch		Y/N	NP (df + N) – S/Pl	-100, 0, +100	Y/Y
	E2b	18	Dutch		Y/N	NP (df + N) – S/Pl	0	Y/N
	E3	26	Dutch		Y/N	NP (df + N) – S/Pl	0	Y/N

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

	Experiment	Participants (n)	Language	Language Family	Agreement context	Type of Phrases	SOA	GCE
	E4a	8	Dutch		Y	NP (adj + N)	-100, 0	N
	E4b	15	Dutch		Y	NP (df + N + adj)	0	Y*
Cubelli et al., (2005)	E1	28	Italian	Romance	N	BN	0	Y
	E2	28			Y	NP (df + N)		N
	E3	28			N	BN		Y
	E4	28			N	BN		Y
Schiller & Caramazza (2006)	E1	28	Dutch	Germanic	Y	NP (df + N)	0	Y
	E2	19						Y
Schiller & Costa (2006)	E1a	20	German	Germanic	N	NP (in + N)	0	N
	E1b	20				NP (df + N)		Y
O'Rourke (2007)	E1	16	Spanish	Romance	N/Y	BN/NP (df + N)	0	N/N
Bordag & Pechmann (2008a)	E1	32	Czech	Slavic	Y	NP (dem + N)	0	Y
	E2	16				NP (adj + N)		
	E3	14				NP (adj + N)		
Heim, Friederici, et al. (2009)	E	14	German	Germanic	Y	NP (df + N)	0	Y
Paolieri, Lotto, et al. (2010)	E1	16	Italian	Romance	N	BN	0	Y
	E2	20	Spanish					Y
Finocchiaro et al. (2011)	E1	24	Italian	Romance	N	BN	0	N
	E2	28	Italian	Romance				N
	E3	30	Spanish	Romance				N
	E4	24	French	Romance				Y
	E5	20	German	Germanic				N
	E6	20	Dutch	Germanic				N
Paolieri et al. (2011)	E1	36	Italian	Romance	N	BN	0	Y
	E2	36						Y
Bürki et al. (2016)	E1	18	German	Germanic	Y	NP (df + N)	0	Y

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Note. Agreement context: Y = Yes (presence of an agreement context); N = No (absence of an agreement context). Type of Phrases: BN = Bare noun; NP = Noun phrase; df = definite article; in = indefinite article; adj = adjective; dem = demonstrative; pron = pronoun; N = Noun; S = Singular; Pl = Plural. SOA = Stimuli Onset Asynchrony. GCE (Gender Congruency Effect): Y = Yes (presence of the GCE); N = No (absence of the GCE); * = significant only by participants or items. Bürki et al. (2016) and O'Rourke (2007) were not included in the analyses due to an insufficient amount of reported statistical information and the lack of raw data.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Table 6

Summary of all the reviewed studies on plural gender-gain effect for the additional meta-analysis

	Experiment	Participants (n)	Language	Nature of agreeing element	Type of Phrases	Nº of stems	Effect
Janssen & Caramazza (2003)	E1	18	Dutch	Freestanding	NP (df + N)	-	Y
Jescheniak et al. (2014)	E	16	Dutch	Bound-morpheme	NP (adj + N)	1	N
Lemhöfer et al. (2006)	E-a	16	Dutch	Free + bound	NP (df + adj + N)	-	Y
	E-b	16		Bound-morpheme	NP (adj + N)	2	Y
Schriefers et al. (2002)	E1	16	German	Freestanding	NP (df + N)	-	Trend
	E3	16		Freestanding	NP (df + N)		Y
Schriefers et al. (2005)	E1	32	German	Free + bound	NP (df + adj + N)	-	Y
	E3	32			NP (adj + N)	2	Y*
Spalek & Schriefers (2005)	E2a-matA	32	Dutch	Freestanding	NP (df + N)	-	N
	E2a-matB	32					Y

Note. Type of Phrases: BN = Bare noun; NP = Noun phrase; df = definite article; adj = adjective; N = Noun; Nº of stems: Number of stems the bound-morpheme can have. Effect: Facilitation effect gained on plural on the gender value with the same agreeing element form in both number values: Y = Yes (presence of effect); N = No (absence of effect); * = significant only by participants or items; Trend = marginally significant.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

5.1.4.1.2 Meta-analytic approach

Our search identified 39 experiments from 14 studies with 75 comparisons of interest for the main meta-analysis of the GCE, and 9 experiments from 6 studies with 10 comparisons of interest for the meta-analysis of the singular plural paradigm. Importantly, one of the requisites of a meta-analysis is that entries (i.e., effect sizes) should be independent. However, we found that multiple comparisons of interest came from the same samples. Thus, we used the robust variance estimation method, which allows for the conducting of a meta-analysis with effect sizes that come from the same sample. The robust variance estimation method is implemented in the *robumeta* package (Fisher et al., 2017) for R (R Core Team, 2018). This package includes a parameter (*r*) to introduce into the software the within-study correlation between effect sizes, and a tool to check the sensitivity of the outcome to different values of *r*. This is accomplished by calculating the meta-analytic estimate when *r* varies in several levels from 0 to 1. In our sample of effect sizes, variations in *r* caused only trivial changes in the results (e.g., changes at the third or fourth decimal) and did not affect the interpretation of the results. We report the results with the default parameter $r = .80$ and we applied the correction for small samples when there were less than 40 participants (Tipton, 2015). Some of the moderators that we tested included more than two levels. In such cases, the *robumeta* package does not report an omnibus test. For those moderators, we report a summary of the pairwise comparisons between levels focusing on significant results. We also conducted comparisons of the meta-analytic estimates against zero, but we only report those that were significant below. From all comparisons, we computed Hedges' *g* as effect size measure. We obtained effect sizes for each comparison of interest between the conditions of gender congruency (GC) and gender incongruency (GI; and for the second and smaller meta-analysis, between singular-plural matching gender value vs. non-matching gender value). Depending on the information reported, we used the following procedures. If *M* and *SD* for both conditions were reported, we used them to compute Cohen's d_{av} (Equation 1):

$$d_{av} = \frac{M_{GI} - M_{GC}}{\sqrt{\frac{SD^2{}_{GI} + SD^2{}_{GC}}{2}}}$$

USC DE SANTIAGO DE COMPOSTELA

where M_{GI} and SD_{GI} are the mean and standard deviation of the response time in gender incongruent trials, and M_{GC} and SD_{GC} are the mean and standard deviation of the response time in the gender congruent trials. Cohen's d_{av} is biased and we applied the correction factor (Equation 2) proposed by Hedges (1981),

$$Hedges'\ g = d_{av} * \left(1 - \frac{3}{4 * df - 1}\right)$$

where *df* is the degrees of freedom of the comparison between GC and GI (*df* = *n* – 1). Finally, *g*'s variance was computed following Borenstein's (2009) Equation 12.21 for matched or paired designs (Equation 3),

$$v_g = \left(\frac{1}{n} + \frac{g^2}{2n}\right) * 2 * (1 - r_{xy})$$

where *n* is the number of participants and r_{xy} is the correlation between response times for GC and GI conditions. This correlation is not usually reported, but we were able to compute it from data sets provided by some authors and estimated the correlation for the other (*r* = .70). If *M* and *SD* were not reported but an *F*-value of the main effect of gender congruency was reported, we followed these steps to compute the effect size. First, we computed partial eta-squared from the *F*-value and its degrees of freedom (Equation 4) following Lakens' (2013) Equation 13:

$$\eta^2_p = \frac{F * df_{\ effect}}{F * df_{\ effect} + df_{\ error}}$$

From η_p we computed Cohen's d_z following Brysbaert (2019) for repeated measures designs (Equation 5):

$$d_z = \sqrt{\left(\frac{\eta^2_p}{1 - \eta^2_p}\right)}$$

Then, we computed d_{av} from d_z (Equation 6) also following Brysbaert (2019):

$$d_{av} = d_z * \sqrt{2 * (1 - r_{xy})}$$

Then, we followed with Equations 2 and 3 above. If, instead of an *F*-value, the report presented a *t*-value, we computed d_z as in Equation 7 in Lakens (2013),

$$d_z = \frac{t}{\sqrt{(n)}}$$

and then proceeded with Equations 6, 2, and 3.

Several comparisons of interest were not significant and the articles only reported *F* < 1, or no inferential statistics at all. In these cases, it was not possible to compute ηp 2 and obtain an effect size. However, removing these non-significant comparisons from the meta-analyses would have overestimated the effects. Thus, although we did not know the actual *F*-value or effect sizes, we estimated them as *g* = 0. This probably caused an underestimation of the meta-analytic estimate, but we found it preferable to the overestimation caused by ignoring non-

USC UNIVERSIDADE DE COMPOSTELA

significant results. In many studies, the same sample provided more than one effect size. To avoid problems derived by the non-independence of the effect sizes, we used the robust variance estimation method (Hedges et al., 2010), which considers the correlation between observations and thus does not require independent effect sizes. All the confidence intervals reported below are 95% *CI*s.

5.1.4.2 Results

5.1.4.2.1 Main meta-analysis: The GCE

We conducted a series of analyses linked with the hypotheses described above. First, we tested the hypothesis that, overall, the GCE is small to moderate. The effect with all the comparisons of interest was indeed small and different from zero, g = 0.131, SE = 0.05, CI [0.029, 0.233], $t(37) = 2.59$, $p = .014$. Heterogeneity between effects was substantial, I^2 = 83.18%, and thus we tried several moderators.

First, we computed the overall effect with German/Slavic languages (k = 56). The effect was different from zero, g = 0.270, SE = 0.044, CI [0.179, 0.362], $t(26) = 6.07$, $p < .001$. For these studies, we compared the effect of the agreement context. The GCE was higher when there was agreement, g = 0.321, SE = 0.050, CI [0.217, 0.425], than when there was not, g = -0.022, SE = 0.063, CI [-0.154, 0.110], $t(21) = 3.75$, $p = .001$. The effect was only different from zero with agreement, $t(21) = 6.44$, $p < .001$. To further explore the conditions that affect the GCE, we selected the effect sizes from studies with German/Slavic languages when there was context agreement (k = 41) and tested the effect of the type of agreement. The GCE was similar between noun phrases with freestanding elements, g = 0.424, SE = 0.061, CI [0.296, 0.553], and noun phrases with freestanding plus bound morphemes, g = 0.425, SE = 0.125, CI [0.163, 0.688], $t(18) = 0.01$, $p = .993$. These two effects were different from zero, $t(18) = 6.93$, $p < .001$ and $t(18) = 3.40$, $p = .003$, respectively. The effects were higher in these two conditions than in the bound morpheme only condition (whose effect was not different from zero), g = 0.042, SE = 0.040, CI [-0.042, 0.125], $t(18) = 5.24$, $p < .001$, and $t(18) = 2.92$, $p = .009$, respectively.

Second, we computed the overall effect with only Romance languages (k = 19). The GCE was negative, g = -0.171, SE = 0.062, CI [-0.309, -0.032], meaning that participants were slower responding to gender congruent items. The effect was different from zero, $t(9.74) = -2.76$, p = .021. If instead of Romance languages we consider transparent languages and thus remove

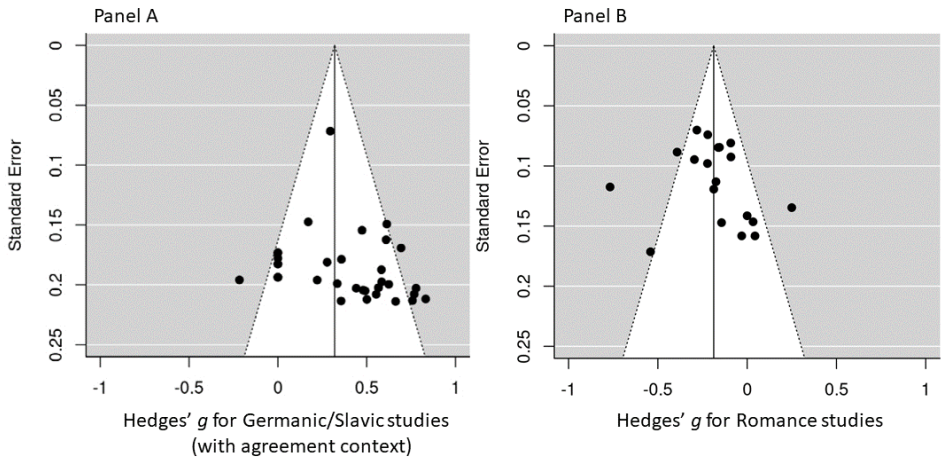
USC UNIVERSIDADE DE SANTIAGO

study 4 from Finocchiaro et al. (2011) with French participants, the effect is still negative, g = -0.128, *SE* = 0.052, *CI* [-0.023, 0-.233], and different from zero $t(36) = 2.47$, $p = .018$.

We also conducted a publication bias analysis to assess the possible overestimation of any effect. We carried out separate analyses per language family. For Germanic/Slavic languages the asymmetry of the funnel plot was significant, $z = 2.33$, $p = .020$. However, assessing the effect this way is not optimal, since it mixes two types of naming instruction together (noun phrases and bare nouns) and hence blurs the effect of determiner congruency (which is the one we are interested in, and differs from a genuine GCE explored with bare nouns). When considering only studies with an agreement context, the asymmetry disappeared: $z = 1.08$, $p = 0.280$ (see Figure 10, Panel A). On the other hand, with Romance languages and bare nouns, the asymmetry is also absent, $z = 0.877$, $p = 0.381$ (see Figure 10, Panel B). The absence of asymmetry is linked to an absence of publication bias, and hence this absence increases the robustness and representativity of the outcomes obtained in our meta-analyses.

Figure 10

Funnel Plots for Studies with Germanic/Slavic Languages (Panel A) and with Romance Languages (Panel B)



5.1.4.2.2 Meta-Analysis on the singular-plural paradigm

In the second meta-analysis, there were only 10 comparisons of interest, and thus the results should be taken with great caution. We found an overall significant effect of facilitation, $g = 0.303$, *SE* = 0.083, *CI* [0.111, 0.495], $t(7.74) = 3.66$, $p = .007$. Heterogeneity was also substantial, $I^2 = 69.98\%$. We examined the effect of the type of agreement and found no differences between conditions, freestanding $g = 0.312$, *SE* = 0.144, *CI* [-0.149, 0.773], morpheme-bound $g = 0.243$, *SE* = 0.097, *CI* [-0.183, 0.669], and both $g = 0.366$, *SE* = 0.276,

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

CI [-3.143, 3.875]. Because of the low number of entries in the meta-analysis, none of the effects were different from zero: freestanding $t(2.98) = 2.17$, $p = .120$, bound morpheme $t(1.95) = 2.51$, $p = .132$, and both $t(1) = 1.32$, $p = .411$. None of the pairwise comparisons were significant: freestanding vs. bound morpheme $t(4.55) = 0.40$, $p = .708$, freestanding vs. both $t(2.33) = 0.17$, $p = .877$, and morpheme-bound vs. both $t(2.34) = 0.42$, $p = .709$.

5.1.4.3 Discussion

The results showed that the gender effect for Germanic and Slavic languages is small but highly significant. When considering the agreement context, we were able to confirm that the "GCE" is consistently present when elements of agreement that vary across gender values are produced, but not when bare nouns or accompanying words that do not vary across gender values are used (e.g., definite articles for nouns in their plural forms in the case of German and Dutch). Overall, the results that consider agreement support the DSIH, that is, the GCE seems to be a determiner congruency effect. Likewise, they are in line with the statement that gender is selected automatically and not through competitive-based mechanisms (Caramazza et al., 2001). Importantly, the size of the gender effects obtained for Romance languages was small, but significant. Hence, support is given to the GIE. This suggests that, at least for Romance languages, nouns compete as a function of gender. Of note, we did not detect any publication bias either in the studies on the "determiner" congruency effect with Germanic and Slavic languages or in the studies on the "gender" effects with Romance languages.

These results allow us to paint the following picture: the PWI paradigm does not in any case reflect effects of gender, but rather competitive effects between lemmas (or lexemes, according to the IN model) in Romance languages and between determiners in Germanic and Slavic languages. The same can be said when adopting the criteria of the overall degree of transparency of the languages rather than of language families (transparent languages [Romance minus French] vs. opaque languages [Germanic/Slavic plus French]). However, there are multiple considerations to be made here. Regarding the GIE, the observation of the effect comes from only one laboratory, which tested native speakers of Italian and Spanish (Cubelli et al., 2005; Paolieri, Lotto, et al., 2010; Paolieri et al., 2011). Only one other laboratory has tried to replicate this with the same languages, and failed to do so (Finocchiaro et al., 2011). In any case, a GIE implies that nouns of the same gender compete, rather than that they are facilitating the selection of the same gender value. As seen in the previous section of this

USC DE SANTIAGO DE COMPOSTELA

chapter, this is hard to interpret theoretically given current theories of language production and the evidence on gender processing. Consequently, the nature of the GIE within the PWI paradigm calls for further research.

Regarding the results with Germanic and Slavic languages, statistical confirmation of the DSIH is obtained, although competition between elements of agreement is the only type of effect that can be identified in these languages through the PWI paradigm, we remain unsure about how this can be taken as evidence for the uncompetitive nature of gender features. In other words, competition between determiners cannot be taken as evidence for the absence of competition between gender features during gender selection. This is especially relevant in this case, as we are restricting ourselves to only one paradigm, which is fairly complex and might be inadequate to detect effects that are too small. Indeed, the PWI paradigm shows mixed results when it comes to detecting smaller effects of form, such as those supposedly arising between bound morphemes, and it does not seem to detect the competition that the selection of the target and distractor's determiners should have according to the Primed Unitized Activation Hypothesis. The GCE for Germanic and Slavic languages (i.e., the determiner congruency effect) that we obtained was indeed significant, but small and heterogeneous. If such an effect between determiners is small and heterogeneous, one might wonder to what extent a supposedly smaller effect involving grammatical features computed before word-form and determiners themselves can arise within this paradigm. In this sense, it is worth recalling here that a) the GIE we detected for Romance languages in the meta-analysis is indeed a very small effect, quite a lot smaller than the determiner congruency effect (in absolute value it would be of .171 vs. the effect of .424 for free-standing elements in Germanic/Slavic languages, which is actually a medium-sized effect), b) studies by Heim, Eickhoff, et al. (2009) and Heim, Friederici, et al. (2009) in German with more fine-grained techniques (fMRI) have found evidence for grammatical gender processing through competitive-based mechanisms.

Importantly, the meta-analysis shows robust evidence for the determiner congruency effect when freestanding elements are present, either alone with the noun or with other words displaying bound morphemes (e.g., definite articles plus adjectives plus nouns). Yet the GCE is not significant for noun phrases with only bound morphemes plus nouns. These results are in line with the ideas of Lapointe and Dell (1989), later defended by Schiller and Caramazza (2003), according to whom bound morphemes are not selected through competitive mechanisms like freestanding elements are, but instead through phonological transformations

USC UNIVERSIDADE DE SANTIAGO

on the stem of the agreeing word. In any case, the number of studies on the issue of bound morphemes is rather small, and so this conclusion must be taken with caution, especially when, once more, we consider the criticisms on the complexity of the PWI paradigm and the possible repercussions it can have when capturing small effects.

Specifically, because of certain insufficiencies of the PWI paradigm, many authors have relied heavily on the singular-plural paradigm (1) to determine whether gender is being selected when plural and diminutive nouns are produced, (2) to assess the mechanisms underlying the selection of bound morphemes. Consequently, we thought that it was crucial to conduct a meta-analysis on the robustness of the effects obtained in this paradigm, as well as to compare their sizes with the effect sizes shown by the PWI paradigm. Certainly, this paradigm suggests that gender is being selected in the plural and diminutive forms and that bound morphemes compete for selection. However, our meta-analysis reports that the number of studies in the literature is insufficient to properly assess the results, obtain effects that are different from zero, and hence to make a proper interpretation of the robustness and significance of these results (only 10 comparisons of interest are available when considering the effect that is usually significant, the effect of facilitation [e.g., for the feminine plural in German, or the common plural in Dutch – note that we cannot mix together the effect regarding number and that regarding diminutive vs. standard forms]). Therefore, we were not able to either confirm that gender is indeed being selected when an agreement context is absent (plural/diminutive forms) or that there is a competitive effect between bound morphemes of different gender in this paradigm. Critically, the results with this task were the main evidence for Schiller and Caramazza's (2003) claim that the GCE is a determiner congruency effect. The authors repeatedly found null effects when using plural definite articles whose only form is "die" in German and "de" in Dutch. The WEAVER++ would say gender was not being selected because agreement is not necessary, and thus, the GCE would not emerge. However, through the singular-plural paradigm the authors confirmed effects of facilitation and cost depending on gender, suggesting that gender was indeed being selected, even in the plural. Thus, the GCE was not a genuine gender effect that depended on the presence of an agreement context, but rather an effect of competition between determiners. As the number of studies on the singular-plural paradigm is not sufficient to establish a significant effect between freestanding determiners, this interpretation of the GCE lacks the degree of robustness that would be desirable.

USC UNIVERSIDADE DE SANTIAGO

In sum, the nature of the statements made in the literature on the GCE with Germanic and Slavic languages is of great relevance for the understanding of language production. These statements involve the mechanisms underlying the selection of grammatical features and morphemes and propose that some of them do not depend on competitive processes. Interestingly, as we have seen, although Schiller and Caramazza (2003; see also Caramazza et al., 2001) repeatedly argue in favour of this kind of mechanism for gender selection, they do not explain precisely how it would work within the architecture of the IN model, which they say fit these results "naturally". Yet Caramazza (1997) sees grammatical features on the syntactic network as nodes that inhibit each other as a result of competition, and so the node that inhibits the other ones is the one supposedly reaching the "threshold" for selection (Caramazza, 1997; p. 195). Therefore, it is not clear how inhibition would occur between two elements that do not interact – compete – with each other. Importantly, these arguments have not been computationally implemented, in contrast to the ideas of WEAVER++. Studies with Romance languages and studies on the singular-plural paradigm might provide a window to test these types of statements. However, there is a notable scarcity of work on grammatical gender either with Romance languages and the PWI paradigm or with the singular-plural paradigm. Thus, it would be interesting to conduct further parametric replications (Regula, 1971)[21] of: a) Cubelli et al.'s (2005), Paolieri, Lotto, et al.'s (2010), and Paolieri et al.'s (2011) studies to determine what is actually happening during gender selection in Romance languages; b) studies on bound morpheme selection, not only with the singular-plural paradigm, but also with the PWI paradigm; and c) studies with Germanic and Slavic languages with more fined-grained measuring techniques such as event-related potentials (Bürki et al., 2016) or fMRI (Heim, Friederici, et al., 2009). Furthermore, it is likewise important to expand the focus of work to other languages, mainly other Romance languages, and to consider the role of other variables in the results, especially the phonological gender transparency of the nouns, which might be a key difference between languages. In our search for answers, we will now address the results on gender processing during language production considering another population, bilingual speakers, and the cross-linguistic GCE.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

5. 2 EVIDENCE ON SECOND LANGUAGE PRODUCTION[22]

[21] We would like to thank Professor Herbert Schriefers for this suggestion.

[22] Reproduced with permission from Springer Nature: Ana Rita Sá Leite, Karlos Luna (*Universidad Nacional de Colombia*), Isabel Fraga, & Montserrat Comesaña (2020). The gender congruency effect across languages in bilinguals: A meta-analysis.

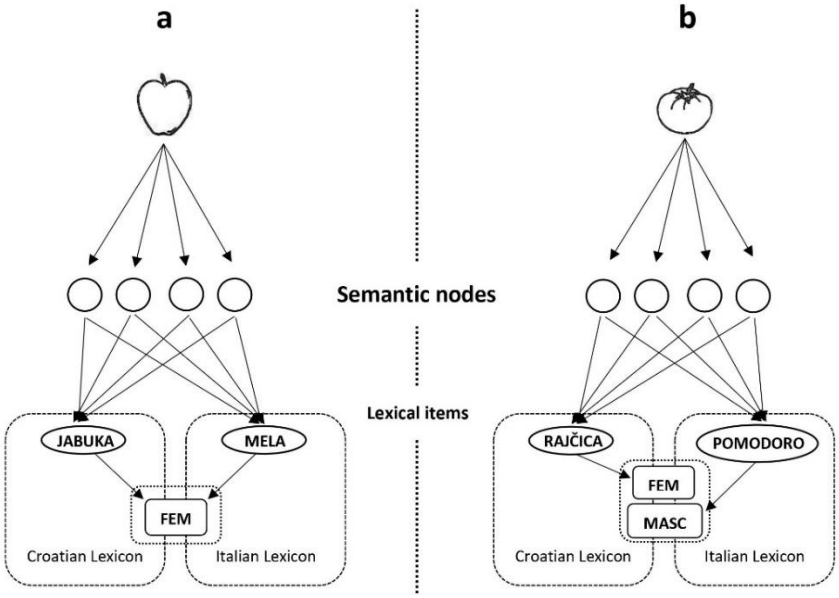
In the bilingual domain, the study of grammatical gender mainly provides evidence for the still unclear debate about the mechanisms underlying its selection. Interestingly, it also allows us to examine the organization of grammatical features across languages. Although types of cross-language interference have been observed at different levels of processing (see Comesaña et al., 2015, 2018; Dijkstra, et al., 2010; Sá-Leite et al., 2019 for more detail; also Soares, Oliveira, et al., 2018; Soares, Oliveira, et al., 2019, for evidence of interactions between lexical-syntactic levels), research at the lemma or grammatical-syntactic stratum is scarcer and more controversial than at the lexeme or morpho-phonological level (see Kroll & Tokowicz, 2005). Hence, over the last decade a number of studies have explored cross-language interactions at the grammatical level through the processing of grammatical gender, testing two conflicting perspectives on the structure of bilingual gender representation: the integrative vs. the autonomous view (Costa, Kovacic, Franck, et al., 2003; Klassen, 2016). Importantly, both views make their predictions on two assumptions: (1) that grammatical gender is an abstract feature subject to competition, and is not accessed automatically, as some authors have argued (Caramazza et al., 2001); and (2) that in line with one of the most influential models of bilingual language organization, the Revised Hierarchical Model (Kroll & Stewart, 1994), there is a direct connection between the two languages at the level of lexical representation (lemma). However, they differ in the organization of gender values across languages in the bilingual mind. According to the integrative view, bilinguals share the representation of grammatical gender values across languages in a unique integrated-gender system. Thus, the activation of a word and its gender value in a specific language affects the gender activation of its translation in the other language (see Figure 11). On the contrary, the autonomous view holds that each language has its own particular gender system. Hence, whether or not two translations share the same gender value is irrelevant for the organization of second language (L2) grammatical knowledge (see Figure 12). Simply put, because gender values are taken to be independent across languages, the gender of a target word does not compete with the gender of its translation equivalent.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Psychonomic Bulletin & Review, *27*(4), 677-693. https://doi.org/10.3758/s13423-019-01702-w (ISSN: 1069-9384; EISSN: 1531-5320).

Figure 11

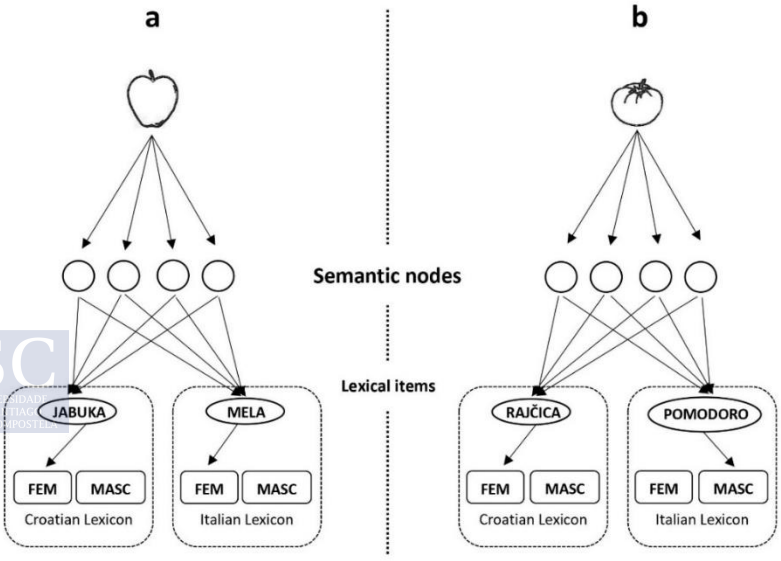
Representation of the gender-integrated hypothesis for words of the same gender across languages (panel A) and those of different gender (panel B)



Note. '*Jabuka*' and '*mela*' mean "apple" in Croatian and Italian, respectively. '*Rajcica*' and '*pomodoro*' mean "tomato". Gender features (feminine [fem] and masculine [masc]) are shared across languages. Figure reproduced from Costa, Kovacic, Franck, et al. (2003) and taken from Sá-Leite et al. (2019).

Figure 12

Representation of the gender-autonomous hypothesis for words of the same gender across languages (panel A) and those of different gender (panel B)



USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Note. '*Jabuka*' and '*mela*' mean "apple" in Croatian and Italian, respectively. '*Rajcica*' and '*pomodoro*' mean "tomato". Gender features (feminine [fem] and masculine [masc]) are independent for both languages. Figure reproduced from Costa, Kovacic, Franck, et al. (2003) and taken from Sá-Leite et al. (2019).

To test the tenets of these two proposals, many authors have based their research on the so-called cross-linguistic GCE. This effect is usually obtained through a simple picture-naming task in which participants have to name aloud a picture in their L2. The automatic activation of the first language (L1) translation equivalents of the L2 nouns that depict the pictures generates the classical gender congruent and incongruent conditions depending on whether or not they share the same gender value. Therefore, conditions are created through the careful selection of homogeneric nouns (i.e., nouns that have the same gender values in both languages) and heterogeneric nouns (i.e., nouns that have one gender value in one language and another gender in the other language). The cross-linguistic GCE consists of faster responses to homogeneric target nouns than to heterogeneric ones. This effect has also been studied in forward translation tasks (i.e., participants are presented with an L1 noun on a screen and they have to translate it into the L2 as fast and accurately as possible).[23]

The first study to examine the cross-linguistic GCE was Costa, Kovacic, Franck, et al. (2003). The authors ran five L2 picture-naming experiments with highly proficient and native-like bilinguals, including different language pairs (i.e., Croatian-Italian, Spanish-Catalan, and Italian-French). Participants were asked to name pictures using a noun phrase formed by a gender-marked definite determiner and a noun (except for the third experiment, in which an adjective following the noun was also included). In their first three experiments, they tested Croatian-Italian bilinguals. Thus, pictures were named using Italian (L2) noun phrases in which the noun's translation could be homogeneric (e.g., '*la mela*' [feminine.] in Italian is '*jabuka*' [feminine] in Croatian, 'the apple' in English [note that in Croatian definite articles do not exist]) or heterogeneric (e.g., '*il pomodoro*' [masculine] in Italian is '*rajcica*' [feminine] in Croatian, 'the tomato' in English). In their fourth and fifth experiments, they tested whether the dissimilarity between gender systems modulated cross-linguistic effects. Thus, they replicated their first experiments, this time with native-like Spanish-Catalan and Italian-French bilinguals, two language pairs that exhibit far greater similarity than the ones featured in their previous

USC UNIVERSIDADE DE SANTIAGO

[23] The cross-linguistic GCE has also been studied in comprehension through lexical decision tasks (i.e., participants are asked to decide whether or not a chain of letters is a real word in a given language; see Lemhöfer et al., 2008). However, the focus of this thesis is language production.

experiments (e.g., whereas Italian, Spanish, Catalan, and French have feminine and masculine gender values, Croatian has a third: the neuter gender value). In all the experiments, monolingual speakers of L2s (Italian, Catalan, or French) were recruited as control groups. The results revealed that whereas no gender effects were found with Croatian-Italian speakers, both the Spanish-Catalan and Italian-French bilinguals showed the effect. However, since the effect was also found with the monolingual control groups, its interpretation as evidence in support of a gender-integrated system was not possible (note that monolinguals should not show the cross-linguistic GCE because they do not activate any translation).

Costa and colleagues explained their findings as either evidence against the integration of both gender systems or as support for the idea of Caramazza and colleagues (2001) that grammatical gender is a feature that cannot be experimentally assessed, since it is not subject to competition. More specifically, if a unique gender system is shared across languages, their experiments should have shown gender effects according to activation-dependent models such as WEAVER++ or even the IN model. First, because the selection of a gender value is achieved by the build-up of activation, which allows competition to take place, and furthermore, because following the requirements of WEAVER++ for gender selection, noun phrases were produced and agreement had to be fulfilled. Conversely, null effects were indeed expected according to an automatic gender-access perspective by which the selection of a noun's lexical node entails the automatic availability of the appropriate gender value for processing (e.g., Caramazza et al., 2001; Schiller & Caramazza, 2003). According to this perspective, gender effects are actually a determiner form-based effect. Because Croatian has no definite articles, the cross-linguistic GCE understood as competition between determiners was absent. Yet this interpretation is questionable. If we assume an integrative view of the bilingual gender system, it is reasonable to think that the Croatian (L1) translation can activate a specific gender node, and that this gender node will activate the corresponding determiner in the L2, creating interference or facilitation for the selection of the correct L2 determiner. Thus, the reason for the absence of results in this Costa's et al. experiment has to lie elsewhere.

In this sense, many criticisms have been raised that may explain these results and weaken the interpretations based on an autonomous view of bilingual gender representation or an automatic selection of grammatical gender. On the one hand, some variables were not properly controlled, such as the phonological gender transparency of the L2 targets (i.e., the number of target nouns that had the typical gender ending letters) or the degree of cognateness of the

USC UNIVERSIDADE DE SANTIAGO

translations (i.e., to what extent L1 translations were not only semantically but also phonologically and orthographically similar to the target L2 nouns; for example, 'lemon': *limón* [Spanish] and *llimona* [Catalonian] vs. 'bed': *cama* [Spanish] and *llit* [Catalonian]). On the other hand, the number of participants was too small in the first three experiments (10 per experiment), resulting in a possible lack of statistical power. Besides, the results might have been influenced by the considerable differences between the gender systems of Italian and Croatian, as well as by the native-like L2 proficiency of participants, a point we will return to below (see Lemhöfer et al., 2008, and Sá-Leite et al., 2019, for an overview).

Several authors have replicated Costa, Kovacic, Franck, et al.'s (2003) study by improving the experimental control and using different sets of stimuli, different pairs of languages, and late-learner bilinguals of multiple degrees of proficiency. Although a cross-linguistic GCE has been systematically observed in the majority of naming studies (Bordag, 2004; Bordag & Pechmann, 2007; Klassen, 2016; Morales et al., 2011; Lemhöfer et al., 2008; Manolescu & Jarema, 2015; Paolieri, Cubelli, et al., 2010) and also in forward translation tasks (from L1 to L2: Manolescu & Jarema, 2015; Paolieri, Cubelli, et al., 2010; Paolieri et al., 2019; Salamoura & Williams, 2007), there are also studies in which the effect has been harder to observe. Significant differences between congruent and incongruent conditions were absent in the analysis of variance by items in many cases (e.g., Klassen, 2016; Lemhöfer et al., 2008; Paolieri, Cubelli, et al., 2010; Paolieri et al., 2019). Besides, there is also an interesting and well conducted study which completely failed to obtain the effect (Bordag & Pechmann, 2008b). The authors conducted three forward translation experiments with Czech-German bilinguals of upper-intermediate to advanced proficiency. In their first two experiments, a condition featuring short (bare nouns) and long (noun phrases: adjective SMALL or BIG + noun) responses was included along with the gender congruency conditions (gender congruent and gender incongruent translations). In the second experiment, a new set of materials was used in which phonological gender transparency was manipulated as a third condition (1/3 of words were transparent, 1/3 opaque, and 1/3 irregular). No cross-linguistic GCE was obtained in either of these experiments. However, in the second experiment transparent nouns were faster and more accurately translated than opaque and irregular nouns, although only in the long condition (noun phrases). The third experiment was conducted because in both the long and short conditions participants only had to translate the noun (i.e., the adjective they had to produce was determined by the size of a "dot" that appeared in front of the word). Consequently, the authors

USC UNIVERSIDADE DE SANTIAGO

wanted to increase the probability of L1 gender retrieval, and included bare nouns but also complex noun phrases to translate (gender marked adjective + noun). Again, no effects were obtained except for the effect of phonological gender transparency for noun phrases in RTs. This was interpreted as evidence that gender in an L2 is computed each time it is needed through different pieces of relevant information, in this case through the phonological form of the word. It could be argued that if the cross-linguistic GCE is an effect of competition between freestanding elements, since bound-morphemes were the only agreement element present, the effects should be null. Nevertheless, this seems unlikely. An inspection of the remaining studies on the matter shows that when bound-morphemes are the only element of agreement present in the noun phrase to be produced, the GCE consistently emerges (Bordag, 2004; Bordag & Pechmann, 2007, Manolescu & Jarema, 2015), including in forward translation tasks (Manolescu & Jarema, 2015; Salamoura & Williams, 2007). See Table 7 for more detail on these studies.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Table 7

Description of the experiments conducted on the GC effect with bilingual population

	Exp.	Task	Part. (n)	L1	L2	L1-L2 Linguistic Family	L2 Transparency	L2 Age of Acquisition (AoA)	L2 Prof.	GC Effect (BNs)	GC Effect (NPs)
Costa, Kovacic, Franck, et al. (2003)	E1	N	10	Croatian	Italian	NR-R	T	Early learners	3	-	Null
	E2		10								
	E3		10								
	E4	N	48	Spanish	Catalan	R-R	T	Early learners	3	-	Y
	E5	N	10	Italian	French	R-R	O	Early learners	3	-	Y
Bordag (2004)	E1	N	18	German	Czech	NR-NR	O	20yo	1	Y	Y
	E2	N	18	Czech	German	NR-NR	O	14yo	2	Y	Y
Bordag & Pechmann (2007)	E1	N	18	Czech	German	NR-NR	O	13.2yo	2	Y	Y
	E2	N	18	Czech	German	NR-NR	O	12.4yo	2	Y	Y
	E3	N	18	Czech	German	NR-NR	O	10.7yo	2	Y	Y
Lemhöfer et al. (2008)	E2	N	16	German	Dutch	NR-NR	O	Late learners	3	Y (-)	Y (-)
	E3	N	22	German	Dutch	NR-NR	O	Late learners	1	Y	Y
Paolieri, Cubelli, et al. (2010)	E1	N	12	Italian	Spanish	R-R	T	21.8yo	2	Y	Y
	E2	N	12	Italian	Spanish	R-R	T	22.2yo	2	Y (-)	-
Morales et al. (2011)	E1	N	32	Italian	Spanish	R-R	T	Late learners	2	Y	-
Manolescu & Jarema (2015)	E1	N	27	Romanian	French	R-R	O	Early learners	3	Y	Y
Klassen (2016)	E	N	17	Spanish	German	R-NR	O	Late learners	1	Y (-)	Y (-)
Salamoura & Williams (2007)	E	T	18	Greek	German	NR-NR	O	10yo	3	N	Y
Bordag & Pechmann (2008b)	E1	T	18	Czech	German	NR-NR	O	12.4yo	2	N	N
	E2	T	18	Czech	German	NR-NR	O	11.8yo	2	N	N
	E3	T	15	Czech	German	NR-NR	O	12.7yo	2	N	N
Paolieri, Cubelli, et al. (2010)	E3	T	12	Italian	Spanish	R-R	T	24.7yo	2	Y	Y
Manolescu & Jarema (2015)	E3	T	27	Romanian	French	R-R	O	Early learners	3	Y	Y
Paolieri et al. (2019)	E1	T	54	Russian	Spanish	NR-R	T	22yo	3	Y (-)	Y (-)
	E2	T	32	Italian	Spanish	R-R	T	28.7yo	3	Y	Y

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Note. Task: N = Picture naming task, T = Forward translation task; L1-L2 Linguistic Family: R = Romance language; NR = Non-Romance language; L2 Transparency = T = Transparent; O = Opaque; L2 Age of Acquisition: yo = years old; When age is not specified the authors did not inform of the exact age of acquisition; L2 Prof. (Proficiency): 1 = Low-Intermediate to upper intermediate; 2 = Intermediate to high proficient, 3 = Advanced/high proficient to nativelike; BNs = Bare nouns; NPs = Noun phrases; GC Effect (BNs) and GC Effect (NPs): Y = Yes (presence of GCE); N = No (absence of GCE); Y (-) = GC effect obtained only in the analysis of variance by subjects; - = condition not tested.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

When carefully reviewing the studies set out in Table 7, we have identified what we believe are the most relevant factors that might be influencing and increasing the slipperiness of the cross-linguistic GCE. The following should be highlighted: (1) the naming instruction (bare nouns vs. noun phrases), (2) the language family or even more specific, the phonological gender transparency of the target language, (3) the L2 proficiency, and (4) the type of task.

The role of the naming instruction in the bilingual literature is quite surprising given the results with the monolingual population, and is closely related to the variable regarding the language family or degree of gender transparency of the target language. As seen, competitive effects of gender are not obtainable within the PWI paradigm for Germanic/Slavic languages (i.e., opaque languages, including French) if an agreement context with variable elements across gender values is absent. This has led Caramazza and colleagues (2001) to claim that grammatical gender, at least in Germanic and Slavic languages, is retrieved automatically without competition between values. Speakers of Romance languages (at least those languages that are transparent, i.e., excluding French), however, show effects in an unexpected direction that is theoretically hard to explain – a GIE – which seems to be reflecting competition between nouns' lemmas on the basis of gender. The picture with regard to bilingualism is rather different. On the one hand, studies seem to observe genuine effects of gender. Indeed, the direction of the effects in bilingualism is always the expected one, at least in theoretical terms regarding grammatical gender: a GCE. The GIE is nowhere to be seen. On the other hand, it seems that the language family is less important when predicting whether these competitive effects of gender between languages emerge or not. In this sense, if gender were indeed retrieved automatically and not through competition based on different degrees of activation, gender competitive effects of any kind would not be obtainable, regardless of the population being assessed, because the core mechanisms of gender selection would not allow this. However, a brief inspection of the literature shows us effects of gender congruency with bare nouns in bilinguals of Germanic and Slavic languages (for instance German and Czech: Bordag 2004; Bordag & Pechmann, 2007, and German and Dutch: Lemhöfer et al., 2008) and Romance languages, regardless of their degree of transparency (for instance Italian and Spanish: Paolieri, Cubelli, et al., 2010; Paolieri et al., 2019; Morales et al., 2011, or Romanian and French: Manolescu & Jarema, 2015). In any case, it is also worth mentioning that, in the third experiment of Lemhöfer et al.'s study (2008) featuring a naming task with L2 Dutch participants, a tendency was observed by which the GCE was greater for noun phrases than for

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

bare nouns. Similarly, Salamoura and Williams (2007), in a forward-translation task with L2 German participants, failed to obtain the effect with bare nouns, but obtained it with noun phrases. In this regard, the variable "phrase type" might play a role in gender processing and must be considered in our current study.[24]

The discussion of phrase type seems closely linked to that of the language family, or perhaps of the degree of gender transparency of the target language. Although insufficient studies are currently available to establish a well-grounded continuum from the most to the least transparent languages, transparency, as we have seen, has a primary role in gender acquisition and has been proposed by Cubelli et al. (2005) as a possible reason for their GIE. Taking into consideration the results on native language production, we must at least contemplate the possibility of a general tendency in the direction of a greater GCE effect for bilinguals of transparent languages (mainly Romance) than for bilinguals of the opaque ones (i.e., Germanic/Slavic languages plus French) when bare nouns are required. After all, those few null results that were obtained in the field are limited to target languages that were opaque, more specifically, from the Slavic or Germanic branches (Bordag & Pechmann, 2008b; Costa, Kovacic, Franck, et al., 2003; Salamoura & Williams, 2007).

Regarding the possible modulations of the GCE due to participants' L2 proficiency, it seems that the higher the L2 proficiency, the shorter the effect. Thus, for instance, Bordag and Pechmann (2007) explain a decrease in the size of the cross-linguistic GCE in their third experiment as a function of the higher L2 proficiency of the participants. Strikingly, in the first three experiments from Costa, Kovacic, Franck, et al.'s (2003) study, conducted with native-like bilinguals (who were also probably simultaneous bilinguals), no cross-linguistic GCEs were found. The proposal that most satisfyingly fits the results from all the reviewed studies is the developmental Bilingual Interactive Activation model (BIA-d, Grainger et al., 2010). The BIA-d is an extension of the connectionist but also localist BIA model (Grainger & Dijkstra, 1992) that aims to explain L2 lexical acquisition and processing. According to this model, both languages of a bilingual are always active to some degree, at least when it comes to late learners. Thus, representations from the non-target language will be active along with the representations of the target language, and thus interaction between both languages will occur. However, the influence of the L1 (non-target language) on the L2 (target language) will be higher when the

[24] Importantly, the discussion on the mechanisms underlying bound morpheme selection in bilingualism does not seem relevant. As seen above, except for Bordag and Pechmann's (2008) study, all studies featuring bound morphemes have found the cross-linguistic GCE, which can be interpreted either as a reflection of gender effects, bound morpheme competition, or both.

proficiency of the bilingual is lower, since during L2 acquisition direct form links between the L1 and their L2 translations are generated. Therefore, bilinguals will rely on the lemma features of the L1 to access L2 lexical entries until they are proficient enough to create their own L2 independent lemmas. Similarly, according to the most recent connectionist model of bilingual language production and comprehension, the Multilink model (Dijkstra et al., 2019), L2 words representations in our lexicon have frequency-dependent resting levels of activation. More specifically, proficiency is closely related to the time of usage of a language, measured through both the age of acquisition of the L2 and its active use (e.g., daily use). Hence, different levels of L2 proficiency are related to different levels of frequency of L2 use and exposure to L2 vocabulary. Although no direct allusions are made to grammatical gender in the Multilink model, following its rationale, it is plausible that the strength of a link between an L2 noun and its grammatical gender value depends on the frequency of use of this noun along with the retrieval of its own gender value.

Finally, the claim that there might be differences between tasks (naming vs. translation) was first made in Bordag and Pechmann (2008b). They pointed out that task requirements might have been responsible for their null results with bilingual of German and Czech. According to their explanation, the time course of the activation of the gender features in L1 and L2 in translation and picture-naming tasks differs. In naming tasks, the activation spreads in parallel, from the concept in common to both L1 and L2 lemmas, to the level of grammatical encoding. Therefore, at the same time as the L1 gender nodes are activated, those of L2 are activated too, and may compete for selection. In forward translation tasks, however, it is the L1 word form and its lemmas that are activated first; the activation then spreads to the lemma of the L2 translation equivalent (or to the concept, and later to the L2 lemma). Consequently, the L2 gender node becomes active after the activation of the L1 gender node. In any case, the GCE has been successfully observed in other studies featuring forward translation tasks, such as Manolescu and Jarema's (2015) second experiment, Paolieri, Cubelli, et al.'s (2010) third experiment, and Paolieri et al. (2019). Hence, the notion of the task being a modulatory variable of the GCE effect is debatable.

USC UNIVERSIDADE DE SANTIAGO

5.2.1 Meta-analysis on the bilingual data

Taking into consideration the literature reviewed thus far, studies on bilingualism suggest that the mechanisms underlying gender selection are indeed based on nodes accumulating

different levels of activation and competing for selection. Indeed, despite the null results observed in Costa, Kovacic, Franck, et al.'s study (2003), many other studies have found evidence for the cross-linguistic GCE. This also gives support to the idea that the languages within the bilingual mind have a gender-integrated system (e.g., Klassen, 2016; Paolieri, Cubelli, et al., 2010). However, the effect in these studies has sometimes not reached statistical significance in analyses of variance by items. Besides, there are many variables, such as task requirements (using bare nouns or noun phrases), the phonological gender transparency of the target language (greatly related to the language family: Romance minus French [transparent] vs. Germanic/Slavic plus French [opaque]), L2 proficiency, and the type of task (naming or translation), that might modulate the effect. The aim of the present study was precisely to examine the robustness of the cross-linguistic GCE and to analyse quantitatively the role of those variables in the stability of the effect. This will shed light on yet unresolved questions as to the nature of grammatical gender representation, as well as to the organization of the bilingual grammatical gender system.

Considering the findings of the studies reviewed, we expected the following results: (1) a small to moderately sized cross-linguistic GCE that supports a gender-integrated perspective; (2) greater cross-linguistic GCEs for bilinguals of transparent L2s in comparison to bilinguals of opaque L2s, especially for bare nouns; (3) a decrease in the size of the cross-linguistic GCE as L2 proficiency rises; and (4) a greater cross-linguistic GCE for naming tasks in comparison to translation tasks.

5.2.1.1 Method

5.2.1.1.1 Literature search

A literature search was conducted on ERIC (Educational Resources Information Center), LLBA (Linguistics and Language Behaviour Abstracts), Psychology Database, PsycINFO, and Google Scholar. Dissertations were not considered. The search included the keywords 'gender congruency' and 'bilingualism' and yielded 362 results. After removing duplicates with RefWorks®, we individually screened all the articles and applied the following criteria for inclusion: (1) papers are on language production, in that the GC effect in comprehension is restricted to only three papers (Lemhöfer et al., 2008 – Experiment 1; Morales et al., 2016; Weber & Paris, 2004) and the cognitive processes involved in production and comprehension, although closely linked, are distinct (see Meyer, 2016); (2) studies explore the

USC UNIVERSIDADE DE SANTIAGO

influence of the L1 on the L2 and not vice-versa – this because the study of how the acquisition of an L2 modifies the processing of the L1 (L2-L1) is based on different mechanisms and principles from those that analyse the influence in the opposite direction (L1-L2, see for instance, Lim & Christianson, 2013); (3) because many studies have found cross-linguistic interaction between an L3 and an L2 during language processing (e.g., Fung & Murphy, 2016; Williams & Hammarberg, 1998), the presence of any L3 might modulate the congruency status of the target, and thus the papers reviewed here test participants that are bilingual speakers of two gendered languages, with no information reported about a third language.

After the application of the inclusion criteria, ten studies were retained as relevant (Bordag & Pechmann, 2007; Bordag & Pechmann, 2008; Costa, Kovacic, Franck, et al., 2003; Klassen, 2016; Lemhöfer et al., 2008; Manolescu & Jarema, 2015; Morales et al., 2011; Paolieri, Cubelli, et al., 2010; Paolieri et al., 2019; Salamoura & Williams, 2007). Citations from these studies were inspected as an additional search, and thus one further study was included (Bordag, 2004). In total, 11 studies were carefully analysed. They most frequently featured participants that were late-learners of an L2 (except Costa, Kovacic, Franck, et al., 2003), thus the conclusions here established must be restricted to bilingual populations that acquired their L2 after approximately the age of 10. Twenty-five experiments were considered and classified according to four different relevant variables, as a means of ensuring comparability in the analysis of the GC effect. First, we considered the naming instruction given to the participants (bare nouns vs. noun phrases). Second, we considered the phonological gender transparency of the target language (the L2) rather than the language family. As specified in the previous section, we classified Italian and Spanish as transparent, and Czech, Dutch, French, and German as opaque. Although we have not assessed Czech before, we would not locate it high on the transparency continuum proposed by Kupisch et al. (2018). Even if we can find some gender cues in this language, they are not as simple as '-o' and '-a' for both grammatical and natural gender. For example, inanimate nouns ending in a consonant tend to be masculine (but they can be feminine if their final consonant is soft or ambivalent), those ending in '-a' and '-e' tend to be feminine, and those ending in '-o' or '-i' are neuter. However, masculine animate nouns with natural gender can also end in '-a'. In addition, not only soft and ambivalent consonants need to be considered (see Naughton, 2005) but also seven cases in which the endings of nouns vary. We are aware that this classification is far from precise, but as seen in Chapter 3, operationalization of the degree of gender transparency has not yet been done in the literature.

USC UNIVERSIDADE DE SANTIAGO

Finally, we divided the experiments according to the L2 proficiency of participants as reported by the authors (intermediate, intermediate to high proficient, high proficient to native-like) and according to the task (naming vs. forward translation tasks). In Table 7 we identify and describe these experiments.

5.2.1.1.2 Meta-analytic Approach

From the 25 experiments, we identified 41 comparisons of interest between gender congruent and gender incongruent trials. Hedges' *g* was computed as measure of effect size as in the previous meta-analysis for each comparison for which enough statistical information was available. For 31 conditions we were able to compute *Hedges' g* from either the published report or from the raw data submitted by the authors. In these cases, we first computed *Cohen's d* following Equations 1, 2, and 3 of the previous meta-analysis. Again, the correlation between response times for GC and GI conditions is not a commonly reported information, but we were able to compute it from data sets provided by some authors. In the 14 comparisons for which correlation was available, it was very high, r = .93, SD = .05, *range* [.80, .99]. Thus, we estimated r = .90 for the other studies.

When published information was not enough, we contacted the authors and requested the original data[25]. Importantly, we obtained the original data from Lemhöfer et al. (2008) Experiments 2 and 3, but although we replicated the published results of Experiment 3, we were unable to do so precisely for Experiment 2. We contacted the first author and tried to understand the discrepancy with her collaboration. We replicated the main pattern of results in the original report, although our computations deviated on average 10 ms from the original. Kristin Lemhöfer kindly agreed that we used the data from our computations in the meta-analysis. Importantly, Bordag (2004) reported two experiments with four comparisons of interest but did not include enough statistical information and thus was excluded. Besides, two experiments from Manolescu and Jarema (2015) were lost from our analysis because they reported the results of both conditions in a collapsed form. More specifically, they did not report statistics separately for each condition and thus only two effect sizes were computed. As both effect sizes included information from bare noun and noun phrase conditions collapsed, they were lost for the meta-analysis with naming instruction as moderator. In addition, the same participants were

[25] We thank Albert Costa†, Kristin Lemhöfer, and Daniela Paolieri for sharing their data. † deceased

used in their Experiments 1 (naming task) and 3 (translation). In the meta-analysis, we identified both experiments with the same label to inform the software that participants were the same. We contacted Bordag and Manulescu and Jarema to request the original data but received no answer. Also, Salamoura and Williams (2005) reported one experiment with two comparisons of interest. They reported statistics for gender congruent and gender incongruent conditions broken down by cognateness (cognate or non-cognate). We contacted the authors, but they were unable to provide a copy of the original data. Thus, we estimated from the published results. For the noun phrases, we started with the reported *F* statistic and computed its square root to get an approximate value of *t*, and then computed Cohen's *d* following Equation 3 in Dunlap et al. (1996). Then, we computed Hedges' *g*. For bare nouns, the exact value of *F* was not reported ($F < 1$). Thus, we averaged *M* and *SD*s from cognate and non-cognate words to get an exact *M* and an approximate *SD* for each gender congruent and gender incongruent trials. This procedure likely overestimated *SDs*, which will result in a lower and more conservative estimation of the effect size. Thus, in total we obtained 35 effect sizes from 22 experiments reported in 10 articles. We computed the effect size from the response time of gender incongruent minus gender congruent cases, that is, we obtained positive *g*'s when response was slower to gender incongruent trials and negative *g*'s when the response was slower to gender congruent trials.

Of note, effect sizes in a meta-analysis have to be independent, but several experiments provided two estimates from the same sample, one for bare nouns and another for noun phrases, and thus some of the effect sizes were correlated. To test the effect of the non-independence of the observations in our results, we conducted the meta-analyses twice, first with a standard approach that considers effect sizes as independent, and then with the robust variance estimation method, developed to take into account the correlation between observations (Hedges et al., 2010). We hence conducted two sets of meta-analyses, one considering the observations as independent and another considering them as correlated. Both used the inverse of the sampling variance to weight observations. For the independent observations meta-analyses, we used the *metafor* package (Viechtbauer, 2010) in R (R Core Team, 2018).

For the correlated observations, we used the *robumeta* package (Fisher et al., 2017) in R (R Core Team, 2018), which implements the robust variance estimation method (Hedges et al., 2010; for an application of this method, see Oswald et al., 2013). To the interested reader, we recommend the excellent vignette prepared by Fisher and Tipton (n.a.) with mathematical explanations and practical examples. The package includes a parameter (*r*) to inform the

USC UNIVERSIDADE DE SANTIAGO

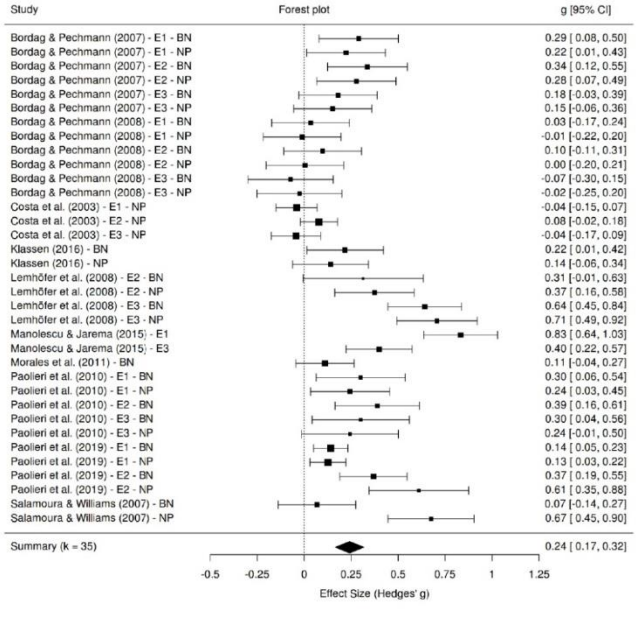
software of the within-study correlation between effect sizes and a tool to check the sensitivity of the outcome to different values of *r*. This is accomplished by calculating the meta-analytic estimate when *r* varies from 0 to 1. In our sample of effect sizes, variations in *r* caused only trivial changes in the results that did not affect interpretation. The outcomes obtained with the *metafor* and the *robumeta* packages were very similar, showing that the within-study correlation between effect sizes did not affect the results and their interpretation. Thus, we report here the results of the meta-analyses with independent observations[26].

5.2.1.2 Results

All the analyses were done in R (R Team, 2018). A series of meta-analyses to test the effect of gender congruency on response time and whether it was moderated by the proposed variables were conducted. In total, 35 effect sizes were obtained and joined the analyses. For the full sample, the effect of gender congruency was small but significant, $g = 0.24$, $SE = 0.04$, 95% *CI* [0.16, 0.32], $z = 6.24$, $p < .001$ (see Figure 13). Heterogeneity was high, $Q_E(34) = 190.27$, $p < .001$, showing a large variability between effect sizes. Several moderators were tested to try to explain this variability, but none showed any difference.

Figure 13

Forest plot with the effect sizes of GC on RTs for all studies with the meta-analytic estimate



[26] The meta-analysis considering correlated observations with the *robumeta* package showed that the effect of gender congruency was small but significant, $g = 0.23$, $SE = 0.04$, 95% *CI* [0.13, 0.32], $t(20.7) = 5.02$, $p < .001$. The moderators did not show any significant differences: for naming instructions $t(19.2) = -0.81$, $p = .43$, for language transparency $t(19.71) = -0.91$, $p = .37$, or for task $t(18.3) = -0.26$, $p = .80$.

Note. BN = Bare Nouns; NP = Noun Phrases.

First, we tested the effect of the naming instructions, bare nouns or noun phrases. Overall, results showed no differences in the cross-linguistic GCE between bare nouns, g = 0.23, SE = 0.05, 95% CI [0.13, 0.33] and noun phrases, g = 0.21, SE = 0.05, 95% CI [0.11, 0.31], $Q_M(1)$ = 0.10, p = .75. We also tested the effect of language transparency depending on whether the L2 of the participants was transparent or opaque. We found no differences between transparent languages, g = 0.20, SE = 0.06, 95% CI [0.08, 0.32], and opaque languages, g = 0.27, SE = 0.05, 95% CI [0.17, 0.36], $Q_M(1)$ = 0.66, p = .42

Linguistic proficiency did not affect the results either, Q_M (2) = 5.04, p = .08, although effect sizes were numerically larger with the lowest proficiency level: for intermediate, g = 0.43, SE = 0.11, 95% CI [0.21, 0.64], for intermediate to high proficiency, g = 0.17, SE = 0.05, 95% CI [0.07, 0.27], and for high proficiency to native-like, g = 0.28, SE = 0.06, 95% CI [0.16, 0.40]. Lastly, with respect to the task that participants completed, naming or translation, no significant differences were found between naming, g = 0.28, SE = 0.05, 95% CI [0.18, 0.38], and translation, g = 0.19, SE = 0.06, 95% CI [0.08, 0.31], Q_M (1) = 1.19, p = .28.

5.2.1.3 Discussion

The main aim of the present meta-analysis was to examine grammatical gender representation and processing in bilinguals in order to shed light on the mechanisms underlying gender selection and, additionally, on the organization of the bilingual mind. To this end, we examined the robustness of the cross-linguistic GCE and we quantitatively assessed the role of certain variables observed in the literature which appeared to be relevant to the stability of this effect. As anticipated, our meta-analysis provided systematic evidence of the cross-linguistic GCE. The effect, though, was small (g = 0.24) and showed considerable variability ($Q_E(34)$ = 190.27, p < .001), which may explain why it is sometimes absent and also why statistical significance is not reached in the analyses of variance by items in a few cases (e.g., Klassen, 2016; Lemhöfer et al., 2008; Paolieri, Cubelli, et al., 2010; Paolieri et al., 2019). Surprisingly, this variability does not seem to be the consequence of the variables that we included in the analysis. The results, hence, go against the automatic gender selection approach (Caramazza et al., 2001) and support the idea that competition between abstract grammatical gender values exists, as held by the GSIH. This because the cross-linguistic GCE does exist and is found to be significant, even in the condition featuring bare nouns. Since the variable “naming

instructions" (bare nouns vs. noun phrases) did not interact with the gender congruency conditions (congruent vs. incongruent), we can conclude that it is found regardless of the production of a bare noun or a noun phrase in bilinguals. In other words, gender effects are indeed observable through experimental assessment and, contrary to the main tenets of the WEAVER++ model (Levelt et al., 1999), an agreement context is not mandatory for the recovery of grammatical gender during lexical access, at least for bilinguals. Thus, competition between different gender values in the lemma (WEAVER++) or syntactic network (IN model, Caramazza, 1997) does occur. Yet this competition is not led by similarity, and hence is not related to competition between lemmas, as the DSM (Cubelli et al., 2005) suggests. The gender effects reported here were of congruency and not incongruency, even for studies with Italian or Spanish as L2 (i.e., the target language). Having established this, the existence of a cross-linguistic GCE supports an integrated view of the bilingual gender system, at least regarding bilinguals who are late learners of an L2. It seems, then, that both languages depend on a shared gender system in which the abstract gender nodes activated by their lexical entries interact during the selection of L2 nouns (see Sá-Leite et al., 2019 for more details).[27]

In this sense, contrary to our predictions, not only the naming instruction (bare noun vs. noun phrase) but also the phonological gender transparency of the L2 (or the language family) and the interaction between both variables failed to modulate the effect. This means that the underlying mechanisms of gender selection somehow change in relation to monolinguals (or at least, native language production) due to certain peculiarities of the bilingual architecture of lexical access. Critically, it is hard for us to believe that these differences between monolinguals and bilinguals imply changes in the core nature of grammatical gender representation. More specifically, we are unsure how the existence of more than one language modifies the "hardware" underlying grammatical gender from "nodes that accumulate activation but do not compete" to "nodes that accumulate activation and compete". In bilinguals, it seems that there is an integrated gender system, similar to the monolingual one but with more lexical entries. These gender nodes compete for selection on the basis of words (translations) of different gender that are connected to them, as we might expect theoretically. In monolinguals or native speakers, we are confused about how and why these gender nodes activated by different lexical

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

[27] Note that there is one study on the cross-linguistic GCE, by Paolieri et al. (2020), that made EEG recordings during language comprehension tasks and found modulations on the N400 component as a consequence of gender congruency. This component is usually related to lexical integration, which the authors interpret in line with the idea of the DSM that grammatical gender is a lexicosemantic characteristic rather than one located in a syntactic network. As we will see in the General Discussion of this thesis, this is highly debatable.

entries of the same language would be automatically retrieved without competition. We propose that it may be the case that grammatical gender selection does indeed always occurs by accumulating activation, bearing competition between gender nodes. However, once again we may be faced with task-related differences and some peculiarities of bilingual language processing that do not change these mechanisms of selection but somehow power the competition that occurs during gender retrieval, making it experimentally observable in bilinguals regardless of the language family or the presence of agreement. In the General Discussion, we will propose a series of reasons that may be behind these differences. Unfortunately, it is difficult for us to understand why we find a GCE and not a GIE when taking into account the results on monolinguals. Yet this GIE that reflects competition between lemmas was theoretically harder to explain in the first place. We will also address this issue in more detail in the General Discussion, after conducting our own experiments on grammatical gender processing.

Moving on to other variables of interest, a consideration of the results of the reviewed studies from a qualitative perspective suggested that the higher the proficiency, the lower the size of the cross-linguistic GCE (e.g., Costa, Kovacic, Franck, et al.'s [2003] null results have been explained through nativelikeness). Modulations on the interaction between languages caused by proficiency are observed widely in bilingual studies (e.g., Grainger et al., 2010; Jackson, 2008). Following the tenets of models such as the BIA-d (Grainger et al., 2010) and the Multilink model (Dijkstra et al., 2019), it is plausible to think that modulations on the cross-linguistic GCE due to L2 proficiency would be expected in one way or another as a consequence of either the development of independent representations of L2 lemmas or variations in the frequency of the use of nouns and their gender values. This would affect the strength of the links between nouns and their grammatical characteristics, and thus would have an impact on the size of the GCE. The absence of modulations of L2 proficiency in the cross-linguistic GCE might be related to the fact that studies differed greatly in terms of how they described proficiency. In fact, this was especially clear when it came to age of acquisition, as well as the years that participants had spent studying the L2 or the time they had been actively using the L2, variables that directly influence (a) the status of the L2 representations in terms of independence regarding the L1 representations, and (2) the frequency of use of nouns and hence "proficiency" as conceived of in the Multilink model. For example, whereas in Bordag and Pechmann's study (2008) intermediate to high-proficiency participants were characterized as

USC UNIVERSIDADE DE SANTIAGO

those who started to learn the L2 after the age of 9 (the age of 14 being the maximum), with an average of 10 years of acquisition, in Paolieri, Cubelli, et al.'s (2010) study, high-proficiency participants were described as those who had started to learn the L2 after the age of 18 and who had been using it for 3.3 years on average. Many seemingly contradictory definitions of high proficiency in different studies can be found (e.g., Paolieri et al., 2019 vs. Salamoura & Williams, 2007). Also, there is an overall lack of objective tests to verify the L2 proficiency of bilingual participants; the self-reported tests seen in many studies also differed in terms of the scale used (e.g., in Paolieri, Cubelli, et al., 2010, a 10-point scale; Lemhöfer's et al., 2008, a 7-point scale). In sum, we should not rule out the possibility of proficiency influencing the GCE, as an imprecise classification of participants in terms of their L2 comprehension and production skills might explain the absence of significant interactions between the GCE and this variable. Future research should address this issue, as it might yield clues as to which mechanisms are behind gender retrieval and what kinds of changes these mechanisms might be experiencing depending on the number of languages sharing the gender system.

Finally, despite a qualitative appraisal of the literature that led us to the provisional suggestion that the type of task (forward translation task vs. naming task) affected the GCE, our analysis failed to confirm this. A tentative explanation of these results can be made based on the variability found in the size of the effect. We might bear in mind here that Bordag and Pechmann (2008b) were the first to suggest this. In their study, the GCE was absent in three forward-translation experiments, and they explained these nulls results in terms of differences in the time course of gender selection between naming and translation tasks, as we noted in the Introduction. However, other researchers have indeed found GCEs using forward-translation tasks (e.g., Manolescu & Jarema, 2015; Paolieri, Cubelli, et al., 2010, 2019; Salamoura & Williams, 2007). Thus, the absence of GCEs in Bordag and Pechmann's experiments might simply be the consequence of the slippery nature of the effect, which is typically small in size and involves a high degree of variability leading to difficulties in its observation (at least, behavioural observation), even in naming tasks (e.g., Costa, Kovacic, Franck, et al., 2003; Klassen, 2016; Lemhöfer et al., 2008; Paolieri, Cubelli, et al., 2010).

USC UNIVERSIDADE DE SANTIAGO

5.3 SUMMARY OF THE MONOLINGUAL AND BILINGUAL DATA

When considering only the monolingual data, the evidence suggests that the GCE for Germanic and Slavic languages has its source in the competition between freestanding elements of

agreement (mainly determiners). Effects with bare nouns or elements of agreement that do not vary across gender values are null. Overall, these results support the DSIH over the GSIH. Yet the GIE obtained with bare nouns for Romance languages suggests that noun lemmas compete on the basis of gender. Regarding the use of noun phrases with speakers of Romance languages, the debate appears a good deal clearer: the systematic absence of effects seems to be a matter of mapping between gender values and the form of the determiners, which causes certain experimental limitations. The results on Romance languages, thus, cannot be interpreted as either positive or negative evidence for the DSIH or the GSIH.

The fact that the GCE seems to be a priori a determiner effect in Germanic and Slavic families and that effects are null when bare nouns are used to name the pictures has been interpreted in favour of the idea that gender is automatically retrieved without competitive-based mechanisms bearing any type of competition for selection (Caramazza et al., 2001). Still, in our opinion, the absence of effects are not sufficient to confirm this claim. Thus, in Germanic and Slavic languages, either (1) gender competition indeed does not occur, or (2) gender competition does occur, but the effect is too small to be captured and is masked within a complex task such as the PWI paradigm. On the other hand, the GIE for Romance languages has been extremely hard to interpret in theoretical terms as competition is occurring between nouns of the same gender rather than between opposite gender values. It is also difficult to understand what the difference between languages is that may cause this situation: Germanic languages do not show competition between lemmas on the basis of gender whereas Romance languages do (in certain studies). Cubelli and colleagues tried to explain this as a function of gender transparency, but their proposal of the DSM based on morphological decomposition had many weaknesses and problematic assumptions. The assessment of the bilingual data uncovers new avenues in the search for possible answers to the unsolved questions raised in the literature on monolinguals. On the one hand, studies with bilingual speakers provide favourable evidence for the presence of competition between gender nodes for every language regardless of its family or, more broadly, its degree of gender transparency, which supports the GSIH. On the other hand, it does so in the theoretically expected direction (i.e., WEAVER++ or IN models), that is, a cross-linguistic GCE (not a GIE), and also regardless of the language family, ultimately contradicting the DSM.

USC UNIVERSIDADE DE SANTIAGO

When comparing the size of the effects, we observed that for monolinguals of Germanic and Slavic languages there is a determiner congruency effect that is medium-sized and

heterogeneous but which is consistently obtained; this represents the largest of the analysed effects. For monolinguals of Romance languages there is a rather smaller GIE that seems to be very slippery, and represents the smallest of the analysed effects. Finally, for bilinguals there is a small cross-linguistic GCE that is almost always obtained but shows more slipperiness than the determiner effect observed for monolinguals of Germanic and Slavic languages. The challenge now is to build bridges between all sets of data to find common ground that can explain the core of grammatical gender representation and selection. However, before moving on to a proposal towards this common ground, we will first present a series of experiments in a Romance and transparent language that has yet to be analysed, European Portuguese. We will subsequently assess the cross-linguistic GCE in European Portuguese-German bilinguals. Our results will be considered along with the reviewed literature when drawing definitive conclusions here.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

6 GRAMMATICAL GENDER IN EUROPEAN PORTUGUESE

In this chapter, we will first describe and analyse the grammatical gender system of European Portuguese, and then present a series of PWI tasks, plus a forward and backward translation task. By using different naming instructions (bare nouns and noun phrases) and assessing different variables that we think may be relevant in the processing of gender within these tasks (mainly animacy, phonological gender transparency, and language proficiency), we intend to:

a) Provide new data on a different transparent Romance language to the pool of evidence with bare nouns, so that we can evaluate the replicability of the GIE shown in Italian and Spanish (Cubelli et al., 2005; Paolieri, Lotto, et al., 2010; Paolieri et al., 2011).

b) Provide new data on a different transparent Romance language to the pool of evidence with noun phrases, ultimately obtaining confirmation of the Late Selection Hypothesis (Miozzo & Caramazza, 1999). As we will explain in detail below, Portuguese is a Romance language that has one-to-one gender-form mapping for determiners, as is also the case with the Germanic/Slavic languages tested so far. The assessment of a language like Portuguese will allow us to confirm the core of the Late Selection Hypothesis. More specifically, the absence of a GCE with noun phrases for Italian and Spanish is understood as a matter of the masking of competition due to the one-to-multiple gender-form mapping of the determiners. This complex mapping between gender and determiners requires that other processes of syllabification and organization within the word phrase take place prior to the selection of the correct determiner. These processes would be masking the competition that occurs between gender features and/or determiners of different gender. If the Late Selection Hypothesis is right and it is indeed a matter of mapping, European Portuguese should show the classical "GCE" as do Germanic and Slavic languages, but if there are other characteristics of Romance

languages behind the absence of effects with noun phrases in these languages, null effects should be obtained with European Portuguese as well.

c) Increase the control of the materials used in previous experiments. As mentioned in Chapter 5, an inspection of the literature reveals that the stimuli used in the PWI paradigm have been poorly controlled in many cases, and multiple variables have never been considered in this field although they are indeed important for picture naming, including, among others, the number of phonological and orthographic neighbours (Grainger, 1988; Zhang et al., 2020). For instance, in the classic studies of Schriefers (1993) and van Berkum (1997) the only controlled variable was the phonological overlap between targets and distractors; Schiller & Caramazza (2003) controlled word frequency, length, phonological overlap and semantic relatedness across conditions; Bordag & Pechmann (2008a) controlled length, semantic relatedness, and transparency across conditions, and others studies such as those conducted by Paolieri and colleagues (Paolieri, Lotto, et al., 2010; Paolieri et al., 2011; also Cubelli et al., 2005) in Italian and Spanish controlled word frequency, length, phonological overlap, transparency, and sometimes semantic relatedness (note that in Paolieri, Lotto, et al., 2010, Experiment 1, they also controlled familiarity). Importantly, in all these studies, the control was always made prior to data collection and the variables were not tested for inclusion in the final models for analysis. Since we are faced here with a complex field with mixed results, we believe it is necessary to guarantee that the basis for experimental assessment is properly controlled. Hence, we intend to conduct experiments that take into consideration the greatest possible number of psycholinguistic variables for the target pictures and nouns as well as distractor nouns. We aim to decrease the differences they may have across conditions, both a priori during stimuli selection and a posteriori through the testing of their inclusion in the models. Among these variables, for pictures we will control the degree of visual complexity and goodness-of-depiction, and for targets and distractor nouns we will consider two measures of per-million frequency (mainly logarithmic frequency), the number of phonological and orthographic neighbours, word length, mean logarithmic bigram frequency, imageability, subjective frequency, and concreteness. The animacy status and phonological gender transparency of targets and distractors must also be either controlled or assessed. Orthographic overlap between target and distractors must be controlled as well, along with first-

USC UNIVERSIDADE DE SANTIAGO

phoneme overlap and semantic overlap. The same variables will be considered in the translation task, including also the degree of phonological overlap between equivalent translations.

d) Among the psycholinguistic variables that must be considered within the PWI paradigm, we detected two that have not been properly assessed in the literature but which may be critical: animacy and phonological gender transparency. Regarding animacy, with the exception of a few studies (Alario & Caramazza, 2002; Bürki et al., 2019; Dank & Deutsch, 2015; Foucart et al., 2010), work within this field has randomly included animate nouns in experimental materials in different proportions. As we have seen, animacy is closely linked to the emergence and organization of gender systems and may be relevant to the way that grammatical gender is processed. Additionally, the animacy status of the stimuli may affect the attentional processes involved in a task such as the PWI paradigm. Regarding phonological gender transparency, little is left to be said about its possible repercussions in language production, especially for transparent languages as conceived of by our systematic procedure described in Chapter 3. Despite this, orthogonal designs considering the possible interaction between grammatical gender and transparency of both targets and distractors processing are scarce. In fact, only Paolieri and colleagues considered this variable in their analyses, but restricted it to target nouns (Paolieri, Lotto, et al., 2010, Experiment 1; Paolieri et al., 2011, Experiment 2). We therefore decided to test not only the gender congruent and incongruent conditions but also the gender of the target noun (masculine vs. feminine) and the transparency category of both targets and distractors (transparent vs. opaque).

e) Once we have described the pattern of results regarding the processing of grammatical gender in native speakers of European Portuguese through the PWI paradigm, we intend to provide data on the cross-linguistic GCE, through a properly controlled sample of early bilinguals of both a transparent and an opaque language (European Portuguese and German). We will focus especially on a variable that we have identified as theoretically relevant but whose impact on the cross-linguistic GCE, according to our meta-analysis, is null: proficiency. As mentioned in the previous chapter, there is a high variability in the way studies describe and assess L2 proficiency. We have detected a need to standardize the assessment of this variable through highly valid and reliable tests, and

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

for this reason, we decided to use the Portuguese and German version of the Dialang (see Alderson, 2006).

In the following pages, we will first describe and assess the gender system of European Portuguese, and then we will present: (a) a set of PWI tasks with bare nouns, to test the replicability of the GIE and the role of animacy in the results; (b) a set of PWI tasks with noun phrases and bare nouns, to test the Late Selection Hypothesis and also the role of phonological gender transparency in the results; and (c) a translation task in both directions with bare nouns, to replicate the cross-linguistic GCE for the first time with European Portuguese-German early bilinguals taking into consideration the role of proficiency.

6.1 THE GRAMMATICAL GENDER SYSTEM OF EUROPEAN PORTUGUESE

European Portuguese is an Indo-European language from the Romance branch that has a bipartite gender system formed by the masculine and the feminine values. Within the animacy hierarchy, we can find nouns with natural gender or with grammatical gender. Most animate nouns have natural gender that can be expressed in the following ways:

a) Through gender inflections (Azeredo, 2013; Câmara Jr., 1970), which are commonly accepted as the expression of the feminine in a new subset of the class represented by the masculine noun by default. The masculine by default may or may not end in the typically masculine "-o": "gato/gat**a**"; "menino/menin**a**"; "professor/professor**a**"; "investigador/investigador**a**".
b) Through derivational gender suffixes for the feminine (mostly ending in "-a" but also in "-iz"): "papa/papisa" (pope/popess); "conde/condessa" (count/countess); "czar/czarina" (czar/czarina); "imperador/imperatriz" (emperor/empress).[28]
c) Although more rare, through phonological alternation and masculine reduction in nouns ending in diphthong: "avô/avó" (/ɐˈvo/, M, grandfather; /ɐˈvɔ/, F, grandmother); "irmão/irmã" (brother, M; sister, F).
d) Through syntactic relationships in common nouns with double gender: "o/a dentista" ("the [M/F] dentist"; "dentista" is a noun with no prior gender value if not expressed through agreement).

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

[28] Some authors consider A and B to be the same case of gender expression: both would be examples of derivation rather than subdividing them into morphemes for gender inflection and derivational gender suffixes (Mathews, 1974; Villalva, 2008). This debate, however, is not relevant for the purposes of the present thesis.

The remaining animate nouns, as well as all inanimate nouns, have grammatical gender, and thus have an intrinsic abstract gender value. The gender of animate nouns with grammatical gender:

a) May coincide with their biological sex: "homem" (M, man), "mulher" (F, woman); "carneiro" (M, ram), "ovelha" (F, sheep).
b) May not coincide with their biological sex (epicene nouns) and further specification of their sex is made through the adjectives "fêmea" (female) and "macho" (male): "águia" (eagle, F), "cangurú" (kangaroo, M), "crocodile" (crocodile, M), "gorilla" (gorilla, M). For instance, "o crocodilo fêmea" (the female crocodile).
c) May be invariant and serve for both sexes: "vítima" (victim, F), "criança" (kid, F).

We will now assess the distributional properties of the European Portuguese gender system. To do so, we will start by analysing the absolute size of each gender value using the Procura-Palavras (P-PAL) database (Soares, Iriarte, et al., 2018). We selected all the available noun forms, in singular, which were nouns (coded as "N" in the category "morf_cat" of the database) besides any other word classes ("governo", which may mean "government" or "to govern", i.e., it can be either a noun or a verb: "N" and "V")[29]. We ended up with 40,739 word forms. These entries include simple and compound nouns, animate and inanimate nouns, and nouns with either grammatical or natural gender, this because the database does not allow us to apply a filter on word complexity, animacy, or type of gender. In any case, our goal is to assess the distribution of gender in the language towards obtaining a picture of the gender environment of the speaker. In this regard, it is useful to understand how much speakers are exposed to one gender value or another, and the fact that all kind of nouns, with both types of gender, formed part of the analysis certainly makes it more informative[30].

Given that nouns naming the letters of the alphabet were present in the database except for the vowels, we added the five vowels, which are masculine. For this, we added the vowels "e", "i", and "u", and classified the entries "a" and "o" – previously identified as nouns but also as "determiners" and/or "prepositions" – only as "nouns". Also, we removed seven entries that were classified in terms of gender as "N/A" because they either showed chains of letters that

USC UNIVERSIDADE DE SANTIAGO

[29] The following combinations were maintained in the final selection: "ADJ, ADJ, N, N", "ADJ, ADV, DET, N", "ADJ, ADV, DET, N, PRONT, QUANT", "ADJ, DET, N, PRON", "ADJ, DET, N, QUANT", "ADJ, INT, N", "ADJ, N", "ADJ, N, QUANT", ADJ, N, QUANT, QUANT", "ADJ, N, QUANT, QUANT, V", "ADJ, N, QUANT, V", "ADJ, N, V", DET, DET, N, PRON, PRON", "DET, N, PRON", "DET, N, PRON, QUANT", "N", "N, ADJ", "N, PRON", "N, QUANT", N, N".

[30] Note that since the analysis is done through word forms and not lemmas, nouns with gender inflections were included as individual entries with their respective characteristics (e.g., "irmã" [sister] and "irmão" [brother], "filha" [daughter] and "filho" [son]).

did not exist ("azanga") or were conjugated verbs ("bafa"). During data treatment we identified another 430 entries that were either foreign words that did not exist in Portuguese dictionaries (e.g., ice-cream), plurals (e.g., "acompanhantes" [companions]), or chains of letters that did not exist at all ("almazém"). This was done by looking up each of these entries in the Portuguese dictionaries *Dicionário Priberam da Língua Portuguesa* (DPLP, Priberam, 2008-2021) and the *Dicionário Porto Editora Infopédia* (Porto Editora, 2003-2021). We divided the remaining 40,306 entries according to gender, thus defining three categories: masculine nouns, feminine nouns, and invariable nouns. The words "coronel", "editorial", "fora-da-lei", "inútil", "nené", "pepel", "skin", "tenente-coronel" and "workaholic" are labelled as "masculine" in the P-PAL. Likewise, the words "doidivanas" and "pop" are labelled as "feminine". However, according to the DPLP and Infopédia dictionaries they have double gender, and hence we moved both entries to the invariable category. Also, the word "TV" was wrongly classified as masculine and thus we corrected it to feminine, as it is listed in the two dictionaries. The words "tira-nódoas" and "dedal" were wrongly classified as feminine and thus were corrected to masculine. From the total of 40,306 entries, we found 2,385 which were classified as having invariable gender, that is, as having double gender (e.g., "ginecologista" [gynaecologist], "pianista" [pianist], "civil" [civilian], "perneta" [one-legged person], "final" [which can be either feminine, "a final" – the final sports match in a competition –, or masculine, "o final" – the end of something –]). Of the remaining entries, 19,422 were classified as masculine (of which one had been left unclassified by the P-PAL and which we classified as masculine, according to the DPLP, "pão-de-forma" [sliced bread]) and 18,499 entries were classified as feminine. This means that in Portuguese, 48.19% of the nouns are masculine, 45.90% are feminine, and the remaining 5.92% have double gender.

We then analysed the entries according to gender transparency. We first adopted the most common definition of phonological gender transparency, considering the nominal endings "o" and "a" as transparent/irregular and the remaining as opaque. We will call this type of transparency "orthographic gender transparency", and we will take into account all types of variations on accentuation within the same category (e.g., category "a" includes endings "-a", "-á", "-ã" and "-â"). This designations means that we will subsequently be able to differentiate it from that of "phonemic gender transparency". The distribution of nominal endings according to gender in terms of orthographic transparency is shown in Table 8.

Table 8

Distribution of the nominal endings in European Portuguese according to orthography

Nominal ending	Masculine	example	Feminine	example	Mostly	TOTAL
-a/-ã/-á	622 (5%)	dia/ecrã/chá [day/screen/tea]	11,684 (95%)	era/maçã/aliá [era/apple/female elephant]	F	12,306
-o/-ô/-ó	12,896 (77%)	número/robô/nó [number/robot/knot]	3,775 (23%)	tribo/câmara-robó/avó [tribe/robot câmera/grandmother]	M	16,671
-b*	1 (50%)	pub [pub]	1 (50%)	web [web]	M/F	2
-c*	2 (100%)	abc [abc]	0 (0%)	-	M	2
-d*	27 (84%)	placard [placard]	5 (16%)	password [password]	M	32
-e/-ê/-é	1,895 (44%)	filme/patê/pé [movie/pate/foot]	2,394 (56%)	parte/mercê/fé [part/grace/faith]	F	4,289
-f*	9 (82%)	surf [surf]	2 (18%)	voz-off [voice-over]	M	11
-g*	70 (97%)	feeling [feeling]	2 (3%)	holding [holding]	M	72
-h*	15 (100%)	brunch [brunch]	0 (0%)	-	M	15
-i	127 (97%)	pai [father]	4 (3%)	lei [law]	M	131
-k*	63 (93%)	rock [rock]	5 (7%)	network [network]	M	68
-l	565 (92%)	lençol [blanket]	46 (8%)	espiral [spiral]	M	611
-m	229 (36%)	jardim [garden]	401 (64%)	imagem [image]	F	630
-n*	93 (99%)	Íman [magnet]	1 (1%)	pen [pendrive]	M	94

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

-p*	24 (92%)	top [top]	2 (8%)	pin-up [pin-up]	M	26
-r	2,203 (98%)	lugar [place]	34 (2%)	cor [colour]	M	2,237
-s	167 (93%)	mês [month]	12 (7%)	vénus [vênus]	M	179
-t*	95 (93%)	copyright [copyright]	7 (7%)	internet [internet]	M	102
-u, -ú	157 (94%)	céu [sky]	10 (6%)	nau [carrack]	M	167
-v*	2 (66%)	leitmotiv [leitmotif]	1 (33%)	tv [TV]	M	3
-w*	6 (100%)	show [show]	0 (0%)	-	M	6
-x*	19 (86%)	fax [fax]	3 (14%)	box [box]	M	22
-y*	44 (96%)	whisky [whisky]	2 (4%)	royalty [royalty]	M	46
-z	91 (46%)	arroz [rice]	108 (54%)	raiz [root]	F	199

Note. Values were calculated using the P-PAL database (Soares, Iriarte, et al., 2018) as described in the text. All nominal endings marked with an asterisk reflect mainly loan words. Translations within brackets.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Considering the total of nouns of the language minus the double-gender nouns, 28,977 of the nouns end either in "-o" or in "-a" or in any of its multiple accentuations, which means that these vowels are the most common endings of the language, being present in 72% of the nouns. When analysing these nouns in terms of transparency, 66% (12,896) of masculine nouns are transparent and 63% (11,684) of feminine nouns are transparent. This means that (1) 61% of the nouns of the language are transparent (24,580 of 40,306), and (2) the masculine gender is slightly more transparent than the feminine gender. Also, 22% of the nouns are opaque, and 11% are irregular. The remaining (6%) have double gender. Until now, we have addressed what would be gender-to-ending predictivity, as defined for French by Colè and colleagues (2003). Looking at the nominal endings, we can address the "ending-to-gender" predictivity. In this sense, 77% of nouns ending in "-o", "-ô", or "-ó" are masculine, and thus are orthographically transparent. On the other hand, 95% of nouns ending in "-a", "-ã", or "-á" are feminine, and thus are orthographically transparent. So, although the masculine gender has more transparent nouns than the feminine gender in both absolute and relative terms, the transparent cue for the masculine, "-o", is far more irregular than the feminine one, "-a". This because 23% of the nouns displaying the masculine cue for gender are feminine, but only 5% of the nouns displaying the feminine cue for gender are masculine.

The rest of the nominal endings would be considered opaque for the reasons explained in Chapter 4, in short, mainly due to: (a) their low presence in the language (approximately 61% of nouns displaying transparent cues vs. 22% displaying opaque cues); that they lack matching with the gender morphemes for agreement and natural gender; and the astonishing amount of evidence showing the role of the transparent cues "-a" and "-o" in gender acquisition and gender retrieval during comprehension. This means that these endings may statistically favour one gender over another, although we will still not consider them transparent. Our analysis shows that many of the opaque endings cover only a few entries ("-b" and "-c" are only present in 2 words each; "-d" in 32; "-f" in 11, and so on). The ending "-e" and its multiple accents is the most common in Portuguese after "-a" and "-o", covering approximately 11% of the nouns. Of these, 56% are feminine, and 44% are masculine, which is indeed a quite balanced situation. The next most frequent endings are "-r" (5.55%), "-m" (1.56%), "-l" (1.52%), "-z" (0.49%), "-s" (0.44%), "-u" (0.41%), and "-i" (0.33%). From these, "-r", "-l", "-s" "-u", and "-i" are mostly masculine (98%; 92%; 93%; 94%; 97%, respectively), and "-m" and "-z" show a more balanced situation favouring the feminine (64% and 56% for the feminine, respectively). Note that the

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

remaining endings are mostly or completely formed by lent nouns (“-b”, “-d”, “-c”, “-f”, “-g”, “-h”, “-k”, “-n”, “-p”, “-t”, “-v”, “-w”, “-x”, and “-y”). In this sense, it is interesting that those endings are mainly masculine (except for “-b”, which shows a complete balance for both endings). The total predominance of the masculine in all endings of lent words coincides with previous evidence regarding the masculine-as-default hypothesis, by which new nouns with no gender arriving in the language are masculine by default (Camara Jr., 1970; Schwindt, 2018).

The analysis based on the view of nominal endings as phonemes (i.e., phonemic gender transparency) gives us a more detailed picture of gender transparency, and is useful for understanding the irregularity observed in the analysis of orthographic gender transparency. The distribution of nominal endings according to gender in terms of phonemic transparency is shown in Table 9.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Table 9

Distribution of the nominal endings in European Portuguese according to phonology

Orthographic representation	Phoneme	Masculine	example	Feminine	example	Mostly	TOTAL
-a	ɐ	555 (5%)	ˈdiɐ (dia) [day]	11,576 (95%)	ˈɛrɐ (era) [era]	F	12,131
-ã, -an (M), -en (M)	ɐ̃	33 (24%)	ɛˈkrɐ̃ (ecrã) [screen]	102 (76%)	mɐˈsɐ̃ (maçã) [apple]	F	135
-á, a (M)*	a	45 (88%)	ˈʃa (chá) [tea]	6 (12%)	ɐˈlja (aliá) [female elephant]	M	51
-o, -u, -ú (M) -out (M)	u	11,847 (98%)	ˈnumәru (número) [number]	262 (2%)	ˈtribu (tribo) [tribe]	M	12,109
-um, -oon (M)	ũ	31 (97%)	ʒәˈʒũ (jejum) [fasting]	1 (3%)	bәˈgũ (begum) [begum]	M	32
-ô, -aux (M)	o	33 (94%)	ʁɔˈbo (robô) [robot]	2 (6%)	ˈkɐmɐrɐ-ʁɔˈbo (câmara-robô) [robot camera]	M	35
-ó, -o (M)*	ɔ	50 (81%)	ˈnɔ (nó) [knot]	12 (19%)	ɐˈvɔ (avó) [grandmother]	M	62
-om (M), -on (M), -d (M)	õ	28 (100%)	ˈsõ (som) [sound]	-	-	M	28
-ão	ɐ̃w̃	1,029 (23%)	kɐrˈvɐ̃w̃ (carvão) [carbon]	3,501 (77%)	kәʃˈtɐ̃w̃ (questão) [question]	F	4,530
-u, -w (M)	w	112 (93%)	ˈsɛw (céu) [sky]	8 (7%)	ˈnaw (nau) [carrack]	M	120
-e, -g, -m (M), -ss (F), -d, -f	(ә)	1,819 (43%)	ˈfiɬm(ә) (filme) [movie]	2,365 (57%)	part(ә) (parte) [part]	F	4,184
-ê, -v (F), -e (M), -d	e	26 (87%)	paˈte (patê) [pate]	4 (13%)	mәrˈse (mercê) [grace]	M	30
-é	ɛ	81 (81%)	ˈpɛ (pé) [foot]	19 (19%)	ˈfɛ (fé) [faith]	M	100
-i, -y, -ie (F)	i	142 (97%)	ˈwiski (whisky) [whisky]	4 (3%)	ˈladi (lady) [lady]	M	146

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

-im, -in (M)	ĩ	116 (99%)	ʒɐrˈdĩ (jardim) [garden]	1 (1%)	ˈbɔlɐ-d(ə)-bərˈlĩ (bola-de-berlim) [berliner (doughnut)]	M	117
-b	b	1 (50%)	ˈpɐb (pub) [pub]	1 (50%)	ˈwɛb (web) [web]	M/F	2
-d	d	20 (87%)	lɛd (led) [led]	3 (13%)	madɐrˈbɔrd (placa-mãe) [motherboard]	M	23
-f, -v (M)	f	10 (91%)	ˈsɐrf (surf) [surf]	1 (9%)	vɔˈzɔf(ə) (voz-off) [voice-over]	M	11
-g	g	62 (98%)	ˈfilĩg (feeling) [feeling]	1 (2%)	ˈołdĩg (holding) [holding]	M	63
-em, -ãe (F)	j̃	59 (13%)	ˈɔmẽj̃ (homem) [man]	409 (87%)	iˈmaʒẽj̃ (imagem) [image]	F	468
-ai (M), -ei, -oi (M), -gh (M)	j	30 (91%)	ˈbɔbslɐj (bobsleigh) [bobsleigh]	3 (9%)	ˈlɐj (lei) [law]	M	33
-k, -c (M), -kh (M)	k	65 (93%)	ˈʁɔk (rock) [rock]	5 (7%)	ˈnɛtwɐrk (network) [network]	M	70
-m	m	11 (92%)	ˈzum (zoom) [zoom]	1 (8%)	ˈsitˈkɔm (sitcom) [sitcom]	M	12
-n	n	70 (99%)	ˈifɛn (hífen) [hyphen]	1 (1%)	ˈpɛn (pen) [pendrive]	M	71
-l	ł	565 (92%)	pɐˈpɛł (papel) [paper]	46 (8%)	(i)ʃpiˈrał (espiral) [spiral]	M	611
-p	p	24 (92%)	piˈnɐp (pin-up) [pin-up]	2 (8%)	ˈtɔp (top) [top]	M	26
-r, -rt (M), -rd (M)	r	2,206 (98%)	luˈgar (lugar) [place]	34 (2%)	ˈkor (cor) [colour]	M	2,240
-t	t	66 (90%)	kɔpiˈʁajt (copyright) [copyright]	7 (10%)	internet (internet) [internet]	M	73
-s, -z, -ch (M), -sh (M), -x (M)	ʃ	268 (69%)	ɐˈʁoʃ (arroz) [rice]	119 (31%)	ʁɐˈiʃ (raiz) [root]	M	387
-x, -ss	s	21 (88%)	ˈfaks (fax) [fax]	3 (12%)	ˈbɔks (box) [box]	M	24

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Note. Values were calculated using the P-PAL database (Soares, Iriarte, et al., 2018) as described in the text. The first orthographic representation is the most common for that phoneme. The phoneme /ɔ/ is represented by "-o" (*) only in the case of the designation of the vowel "o". The phoneme /a/ is represented by "-a" (*) only in the case of the designation of the vowel "a". Translations in brackets. There are words that are counted in different phonemic categories in the masculine categories ("bouquet", "flux", and "futon"), so that the final count is three values higher than the 19,422 masculine entries we have.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

The most frequent phoneme for the masculine is /u/ (represented mostly by letters "-o" and "-u"), being present in 61% of the masculine nouns; the most frequent phoneme for the feminine is /ɐ/ (represented by letter "-a"), being present in 63% of the feminine nouns. In terms of "ending-to-gender" regularity, the phoneme "u" is masculine in 98% of the cases, but /ɐ/ is feminine in 95% of the cases, showing slightly higher levels of irregularity and reversing the values obtained for the previous orthographic analysis. This inversion, as well as the higher levels of irregularity observed for the feminine when analysing orthographic gender transparency (more feminine nouns ended in "-o" than masculine nouns ended in "-a", which was a less irregular ending than "-o") is mainly due to the phoneme /w̃/, represented by "-ão" and being left out of the count for masculine transparency with the phoneme /u/. More specifically, 77% of the nouns (3,501 entries) ending in this phoneme are feminine and only 23% are masculine (1,029 entries). Similarly, although not as significant, since it only covers 51 entries, one of the endings that we had orthographically labelled as transparent for the feminine value is actually masculine in 88% of the entries: "-á" (chá [tea]), representing the phoneme /a/. It is also interesting to note that most phonemic nominal endings are masculine (except for /ʃ/ and /(ə)/), mirroring the situation seen for orthographic transparency. Statistical balance is attained by phoneme /b/ (50%, which covers only 2 entries) followed by phonemes /(ə)/ (43% [M] vs. 57% [F]) and /ʃ/ (69% [M] vs. 31% [F]).

Although a complete overview of the language is interesting and broadly informative, we must consider what would happen if we were to account for the relative distribution of each gender. In this way we can better understand the gender environment of the speaker, getting a more precise picture of how often masculine or feminine nouns are encountered and how often these are transparent, opaque, or irregular. To assess this question, we used the SUBTLEX-PT database (Soares et al., 2015), since it has been shown that subtitles-based databases are more precise and show a greater degree of reliability when representing the real frequency of exposure to words within a language (Brysbaert & New, 2009; Dimitropoulou et al., 2010; Keuleers et al., 2010; New et al., 2007). We therefore selected the 500 most frequent nouns[31] of European Portuguese, and only kept those in their singular form (68 plural nouns were removed). The nouns "equipe" (team) and "mamãe" (mommy) were discarded since they

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

[31] Note: the following classes were selected: "ADJ, V, N", "N", "N, ADJ", "N, ADJ, ADV", "N, ADJ, QUANT", "N, ADJ, V", "N, ADV", "N, ADV, V", "N, CONJ, V", "N, INT", "N, INT, V", "N, REP", "N, V", "N, V, ADJ", "N, V, ADJ, ADV", "N, V, ADJ, INT", "N, V, ADV, ADJ", "N, V, INT", "N, V, INT, ADJ".

belong to Brazilian Portuguese and are barely used in European Portuguese, as reported in the DPLP and Infopédia dictionaries. A total of 28 nouns were classified as invariable for gender (they were double-gender nouns) and were removed. A final sample of 402 of the most common nouns according to the SUBTLEX-PT remained in the list. Note that among them were nouns with either natural or grammatical gender. The analysis of these entries allowed us to confirm that most nouns are indeed masculine, most nouns are transparent, and the feminine value does include more irregular endings ("-o") than the masculine one. Interestingly, the number of transparent nouns is similar for both genders. Hence, of these 402 entries, 52% (211) are masculine and 48% (191) are feminine. Descriptions of the nominal endings according to orthographic and phonemic transparency are shown in Tables 10 and 11, respectively.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Table 10

Distribution of the nominal endings according to orthography for the most frequent nouns of European Portuguese

Nominal ending	Masculine	Example	Feminine	example	Mostly	TOTAL
-a/-ã/-á	6 (5%)	dia/papá [day/father]	126 (95%)	coisa/irmã [thing/sister]	F	132
-o/-ô/-ó	144 (87%)	mundo/avô [world/grandfather]	21 (13%)	mão/ avó [hand/grandmother]	M	166
-b*	0	-	0	-	-	0
-c	0	-	0	-	-	0
-d*	0	-	0	-	-	0
-e/-é	21 (39%)	filme/pé [movie/foot]	33 (61%)	noite/sé [night/cathedral]	F	54
-f*	0	-	0	-	-	0
-g*	0	-	0	-	-	0
-h*	0	-	0	-	-	0
-i	3 (75%)	pai [father]	1 (25%)	lei [law]	M	4
-k*	0	-	0	-	-	0
-l	9 (100%)	-	0 (0%)	-	M	11
-m	4 (57%)	fim [end]	3 (43%)	viagem [trip]	M	7
-n*	0	-	0	-	-	0
-p*	0	-	0	-	-	0
-r	15 (88%)	lugar [place]	2 (12%)	dor [pain]	M	17
-s	5 (83%)	mês [month]	1 (17%)	miss [miss]	M	6
-t*	0	-	0	-	-	0
-u, ú	2 (100%)	céu [sky]	0 (0%)	-	M	2

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

-v*	0	-	0	-	-	0
-w	0	-	0	-	-	0
-x*	0	-	0	-	-	0
-y*	0	-	0	-	-	0
-z	2 (33%)	nariz [nose]	4 (67%)	luz [light]	F	6

Note. Values calculated using the SUBTLEX-PT database (Soares et al., 2015). All nominal endings marked with an asterisk reflect mainly loan words. Translations in brackets.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Table 11

Distribution of the nominal endings according to phonology for the most frequent nouns of European Portuguese

Orthographic representation	Phoneme	Masculine	example	Feminine	example	Mostly	TOTAL
-a	ɐ	5 (3%)	plɐˈnɛtɐ (planeta) [planet]	123 (97%)	ˈkazɐ (casa) [house]	F	127
-ã	ɐ̃	0	-	3 (100%)	mɐˈmɐ̃ (mamã) [mom]	F	3
-á	a	1 (100%)	pɐˈpa (papá) [dad]	0 (0%)	-	M	1
-o, -u	u	137 (100%)	ˈmũdu (mundo) [world]	0 (0%)	-	M	137
-ô	o	1 (100%)	ɐˈvo (avô) [grandfather]	0 (0%)	-	M	1
-ó	ɔ	0	-	1 (100%)	ɐˈvɔ (avó) [grandmother]	**F**	1
-om	õ	2 (100%)	ˈsõ (som) [sound]	0 (0%)	-	M	2
-ão	ɐ̃w̃	7 (26%)	irˈmɐ̃w̃ (irmão) [brother]	20 (74%)	ˈmɐ̃w̃ (mão) [hand]	F	27
-u	w	1 (100%)	ˈsɛw (céu) [sky]	0 (0%)	-	M	1
-e, -m (M), -ss (F)	(ə)	19 (38%)	ˈfiɬm(ə) (filme) [movie]	31 (62%)	ˈnojt(ə) (noite) [night]	F	50
-é	ɛ	2 (67%)	ˈpɛ (pé) [foot]	1 (33%)	ˈsɛ (sé) [cathedral]	M	3
-i (M), -ie (F)	i	1 (50%)	ˈmi (mi) [mi]	1 (50%)	(i)ʃˈpɛsji (espécie) [species]	M/F	2
-im	ĩ	1 (100%)	ˈfĩ (fim) [end]	0 (0%)	-	M	1
-em, -ãe (F)	ɐ̃j̃	1 (20%)	ˈɔmɐ̃j̃ (homem) [man]	4 (80%)	ˈmɐ̃j̃ (mãe) [mother]	F	5

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

-ai (M), -ei	j	2 (67%)	ˈpaj (pai) [father]	1 (33%)	ˈlɐj (lei) [law]	M	3
-l	ɫ	9 (100%)	pɐˈpɛɫ (papel) [paper]	0 (0%)	-	M	9
-r	r	15 (88%)	ɐˈmor (amor) [love]	2 (12%)	muˈʎɛr (mulher) [woman]	M	17
-z, -s (M)	ʃ	7 (64%)	nɐˈriʃ (nariz) [nose]	4 (36%)	ˈluʃ (luz) [light]	M	11

Note. Values calculated using the SUBTLEX-PT database (Soares et al., 2015). The first orthographic representation is the most common for each phoneme. The phoneme ɔ is represented by "-o" (*) only in the case of the designation of the vowel "o". Translations in brackets.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

In terms of orthographic transparency, 67% (144) of the masculine nouns are orthographically transparent (ending in "-o" or "-ô"), 30% (61) are opaque, and 3% (6) are irregular (ending in "-a" or "-á"). Regarding feminine nouns, 66% (126) are orthographically transparent (ending in "-a" or "-ã"), 23% (44) are opaque, and 11% (21) are irregular. From the perspective of ending-to-gender regularity, the nominal ending "-o" (and its variations) again shows most irregular cases: 13% of the entries with this ending are feminine, but only 5% of the entries with the nominal ending "-a" (and its variations) are masculine. Among the opaque endings, "-e" (including "-é") is again the most common ending, being present in 13% of the nouns, of which 61% are feminine and 39% are masculine, a more unbalanced situation than the one seen when we considered the absolute proportion of transparency. The ending "-e" is followed by the endings "-r" (4.23%), "-l" (2.74%), "m" (1.74%), "s" and "z" (both 1.49%), and finally "-i" (1%) and "-u" (including the stressed "-ú", 0.5%). These endings are all mainly masculine, except for the case of "-z" which is feminine in 67% of the cases, an exception that had already been seen in the first analysis regarding the absolute distribution of transparency.

Finally, in terms of phonemic transparency, as can be seen in Table 11, one phoneme changes the proportional pattern between genders in comparison to the phonemic analysis of the absolute distribution: /ɔ/ switches from the masculine to the feminine, as it is only present in one noun, which happens to be feminine: "avó" (grandmother). Most phonemic nominal endings are again masculine, and the feminine transparent phoneme /ɐ/ is again more irregular than the masculine /u/. The phoneme /u/ is present only in masculine nouns, but the phoneme /ɐ/ has a few exceptions, with 5% of nouns with this ending being masculine (e.g., "planeta" [planet]). These two phonemes, however, account for 67% of the entries here. After these, the most frequent phonemes are: /(ə)/ (12% of the sample); /w̃/ (7%) represented by "-ão", and which is mainly feminine (74%); and /r/ (4%). The ending phonemes showing the most balanced proportion between genders are /(ə)/ (38% [M] and 62% [F]), /ʃ/ (64% [M] and 36% [F]), and /i/ (50%), although the latter is only present in 2 entries.

To conclude, we note that transparency and gender can also be analysed in relation to derivational suffixes. However, as seen in Chapter 3, derivational suffixes do not comply with the criteria of monofunctionality and are not the basis of gender transparency, since gender acquisition in transparent languages is not shaped by them. However, because they are of interest in understanding how certain orthographic nominal endings (such as "-e") are mainly of one gender value or another, depending on the suffix to which they belong, we will address

USC UNIVERSIDADE DE SANTIAGO

them below. Certainly, the relation between derivational suffixes and gender has already been described widely in the literature (e.g., Said Ali, 1971). In any case, in Table 12 we have classified the European Portuguese derivational suffixes according to their gender, based on Rio-Torto et al.'s (2013) analysis of nouns formation through suffixation.

Table 12

List of European Portuguese derivational suffixes according to gender

Type of suffixation	Derivational Suffix	Example	Gender
Deadjectival nouns	-eira	asneira [silly thing; mistake]	Feminine
	-ez/a	altivez [arrogance]	Feminine
	-ia	alegria [happiness]	Feminine
	-ice	chatice [nuisance]	Feminine
	-idade	agilidade [agility]	Feminine
	-idão	aptidão [aptitude]	Feminine
	-ura	doçura [sweetness]	Feminine
	-ismo	fatalism [fatalism]	Masculine
Denominal nouns	-ada	facada [stab]	Feminine
	-agem	folhagem [greenery]	Feminine
	-ama	mourama [Moorism]	Feminine
	-aria	papelaria [stationery shop]	Feminine
	-eira	fruteira [fruit bowl]	Feminine
	-ia	capitania [captaincy]	Feminine
	-ina	cafeína [caffeine]	Feminine
	-ite	amigdalite [tonsillitis]	Feminine
	-ose	fibrose [fibrosis]	Feminine

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

	-ado	consulado [consulate]	Masculine
	-al	areal [sandpit]	Masculine
	-ame	velame [canopy]	Masculine
	-ário	preçário [price list]	Masculine
	-ato	colonato [land tilled by settlers]	Masculine
	-edo	arvoredo [groove of trees]	Masculine
	-eto	carboneto [carbide]	Masculine
	-eiro	saleiro [salt shaker]	Masculine
	-ês	economês [economese]	Masculine
	-il	canil [dog kennel]	Masculine
	-io	mulherio [women]	Masculine
	-ismo	darwinismo [Darwinism]	Masculine
	-ito	clorito [chlorite]	Masculine
	-oma	fibroma [fibrous tumor]	Masculine
Event deverbal nouns	-agem	filtragem [filtering]	Feminine
	-aria	destilaria [distillery]	Feminine
	-ção	contaminação [pollution]	Feminine
	-da	balada [ballad]	Feminine
	-deira	brincadeira [joke, children's play]	Feminine
	-dela	telefonadela [call – colloquial form]	Feminine
	-dura	cercadura [cincture]	Feminine
	-eira	quebreira [fatigue]	Feminine
	-ice	aldrabice	Feminine

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

		[fraud]	
	-nça	vingança [vengeance]	Feminine
	-ncia	falência [insolvency]	Feminine
	-ão	empurrão [push]	Masculine
	-ço	cansaço [fatigue]	Masculine
	-ido	zumbido [buzzing]	Masculine
	-mento	desaparecimento [disappearance]	Masculine
	-nço	falhanço [failure]	Masculine
	-tório	interrogatório [interrogatory]	Masculine
Individual deverbal nouns	-dora	escavadora [bulldozer]	Feminine
	-doura	manjedoura [crib]	Feminine
	-deira	lançadeira [shuttle]	Feminine
	-ória	convocatória [summon, convocation]	Feminine
	-eira	coalheira [rennet]	Feminine
	-eta	chupeta [pacifier]	Feminine
	-alha	acendalha [tinderbox]	Feminine
	-ilha	cortilha [dough scrapper]	Feminine
	-iva	iniciativa [initiative]	Feminine
	-dor	amortecedor [shock absorber]	Masculine
	-nte	detergente [detergent]	Masculine (or double gender)
	-douro	bebedouro [drinking fountain]	Masculine
	-ão	espião [male spy]	Masculine

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

	-vel	dirigível [steerable]	Mostly masculine
	-deiro	desfiladeiro [ravine]	Masculine
	-ório	conservatório [conservatory]	Masculine
	-al	estendal [dryer]	Masculine
	-eiro	calceteiro [paver]	Masculine
	-alho	espantalho [scarecrow]	Masculine
	-elho	rapelho [earwig]	Masculine
	-ilho	atilho [tie]	Masculine
	-ulho	tapulho [plug]	Masculine
	-ivo	comprovativo [receipt, proof]	Masculine
	-ário	enxertário [parrel]	Masculine

Note. The list follows Rio-Torto et al.'s (2013) analysis of noun formation through suffixation. Deadjectival nouns are formed from adjectives and denote properties, qualities, or characteristics of the adjectival stem and/or the state, attitudes, or entities characterized by this property. Denominal nouns are formed from other nouns and are related to the following semantic classes: event, individuals, local, cause/source, and result/product. Deverbal nouns are formed from verbs and are related to two large semantic classes of events and individuals. We have not included nouns converted from adjectives (e.g., "cervejeiras" [beer factories]; "preguiçoso" [lazy]; note that whereas deadjectival nouns are created by adding suffixes to adjectives, nouns converted from adjectives maintain the form of the adjectives (that can have suffixes of their own) without any additional suffix – it is said that the adjective, by usage, has been nominalized; for more details, see Rio-Torto et al. [2013; p. 173]). Although sometimes the suffixes display the same combination of letters, they serve different derivational purposes, and hence are considered different suffixes (e.g., "-agem" for denominal nouns vs. "-agem" for event deverbal nouns).

Gender is distributed equally (50% per value) among the 72 derivational suffixes identified by Rio-Torto et al. (2013). Of the 36 feminine suffixes, 26 end in the

orthographically/phonologically transparent vowel "-a"; of the 36 masculine suffixes, 26 end in the orthographically/phonologically transparent vowel "-o". Note interesting cases such as "-ão", a combination of letters that is used in two different suffixes that form deverbal nouns related to either individuals or events. It is always masculine and must be distinguished from the suffixes "-ção" (to form event deverbal nouns) and "-idão" (to form deadjectival nouns), which are always feminine. We can also find another case, "-oma", in which the derivational suffix includes the orthographic nominal ending of the opposite gender in terms of orthographic gender transparency ("-a", but is masculine). Finally, the typical ortho-phonologically opaque ending "-e" is always masculine when belonging to suffixes "-nte" (to form deverbal nouns), and "-ame" (to form denominal nouns), but feminine when belonging to suffixes "-ice" (to form deverbal nouns), "-ite" and "-ose" (to form denominal nouns), and "-ice" and "-idade" (to form deadjectival nouns).

The description and analysis of the European Portuguese gender system allows us to operationalize its degree of gender transparency for our purposes, based on the degree of the relative complexity of gender systems as defined by Audring (2019). In Table 13, we have detailed such an operationalization. The only element missing is the frequency of exposure and use of transparent cues for gender during the first years of life. Current databases of language development in European Portuguese (e.g., FrePOP, Frota et al., 2010; PLEX5, Frota et al., 2012; CDI-European Portuguese, Frota et al., 2015; ESCOLEX, Soares et al., 2014) do not allow us to conduct such an analysis, and future research might usefully fill this gap.

Table 13

Operationalization of the degree of gender transparency of European Portuguese

Factor	Modality	Subfactors	European Portuguese
Frequency	**Phonology**	Proportion of gender cues in nouns	60%
		Number of gender cues per value	61% (M) and 63% (F)
	Morphology	Proportion of gender morphemes in nouns and agreeing elements	Highest
		Number of gender morphemes per value	Lowest (one per gender value)
	Both	Frequency of exposure and use during childhood	- unavailable
		Overall frequency of exposure and use	Highest possible number of nouns displaying those cues
		Number of gender values	2
Perspicuity	**Phonology**	Simplicity	1 cue per gender
		Syncretism	Ending-to-gender predictability equals gender-to-ending predictability

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

		Matching between phonology and orthography	High
	Morphology	Simplicity	Simplest
	Both	Syncretism (in agreeing elements)	Agreeing elements represent only one value
Consistency	**Phonology**	Proportion of irregular nouns	Lowest
		Proportion of opaque nouns	Lowest
		Mapping of gender values to gender cues	One-to-one
	Morphology	Coincidence between phonological gender cues and gender morphemes	Absolute for the feminine; high for the masculine
Monofunctionality	**Phonology**	Intertwining with other features	Null
		Gender cues are monofunctional	Cues are not derivational suffixes

Indeed, European Portuguese is a language that displays a relatively high number (+60%) of simple gender cues ("a" vs. "o" or /ɔ/ vs. /u/) for each value, with the minimum number possible of gender values to be acquired (2). Besides, it shows an almost absolute match between ortho-phonological gender cues and gender morphemes for the feminine ("-a"; an exception is the derivational feminine morpheme "-iz") and a very substantial match for the masculine (morpheme "-o" [/u/] and ortho-phonological cue "-o" [/u/]), as well as pure monofunctionality (i.e., these transparent cues of gender have null intertwining with other features [e.g., number, declension, or case] and null polyfunctionality of the gender nominal endings [e.g., "-a" and "-o" are not derivational suffixes]). However, Portuguese is not a highly orthographically transparent language, and so there is a certain degree of mismatch between orthographic forms and phonemes. This leads to situations in which accentuated forms of the nominal endings that rely on the typical vowel for each gender ("-o" and "-a") actually correspond to other phonemes that are not the typical phoneme for a specific gender value, leading to a certain degree of irregularity in orthographic terms. That is the case, for instance, with the orthographic nominal ending "-á", which is a typical ending for the masculine value in phonological terms (phoneme /a/) despite including the letter "-a". In any case, the irregularity seen in European Portuguese is not comparable to that of other languages, such as French. Indeed, we have not identified any type of ending-to-gender type of irregularity (as "-e" in French) that may cause confusion in terms of gender-to-ending transparency.

Although it would be ideal to calculate the degree of gender transparency of European Portuguese, the first step to do so is to compare it to other languages whose degree of transparency is operationalized following the same standardized method. In that way, we would be able to locate it on the transparency continuum proposed by Kupisch et al. (2018). Since

none of the languages seen on this continuum have had their degree of morpho-phonological gender transparency systematically assessed through linguistic descriptions and word databases, we cannot take a position as to any precise conclusions regarding the degree of gender transparency of European Portuguese. Yet we can certainly affirm that Portuguese would be on the transparent side of the continuum. Furthermore, a brief inspection of Spanish and Italian allows us to hypothesize that they would be more gender transparent than European Portuguese. In Spanish, whereas the distribution of phonemic and morphemic endings across genders is very similar to that of European Portuguese, there is substantial ortho-phonemic consistency (i.e., it is a highly orthographically transparent language; Goswami et al., 1998; Seymour et al., 2003). Although European Portuguese is also considered an orthographically transparent language (e.g., Girolami-Boulinier & Pinto, 1994; Viana et al., 1991), it has less consistent and more complex grapheme-phoneme mappings than Spanish (Defior et al., 2002). Hence, in Spanish, there is no mismatch between phonological and orthographical gender transparency (ending "-a" always refers to phoneme "/a/"; ending "-o" always refers to phoneme "/o/"). Even though in Portuguese this mismatch is not considerable (as it is in French, for instance), it is greater than in Spanish. The same applies to Italian, which also relies even more on the "-a" and "-o" endings, having a smaller set of opaque endings (more often than not "-e"). Hence, we believe that when locating these three gender transparent languages on a continuum, they are extremely similar, but with European Portuguese exhibiting a lower degree of gender transparency than Spanish and Italian.

6.2 Study I: Bare nouns and animacy[32]

In the present study, we are interested in the representation and processing of grammatical gender in language production in the specific case of European Portuguese. As seen in Chapter 5, research on grammatical gender retrieval within the PWI paradigm in Romance languages is scarce and reveals a complex situation of mixed results. We conducted a series of meta-analyses in an attempt to disentangle this situation and observed a very small but significant GIE. This suggests that gender representation and processing may be closely tied to the characteristics of languages themselves. More specifically, at least in Romance languages (or transparent languages), it seems that lemmas compete on the basis of gender. This contrasts with the picture

[32] Based on the paper: Ana Rita Sá Leite, Juan Haro (*Universitat Rovira I Virgili*), Montserrat Comesaña, & Isabel Fraga (2021). Of beavers and tables: The role of animacy in the processing of grammatical gender within a picture-word interference task. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.66117 (ISSN: 1664-1078; EISSN: 1664-1078).

drawn for Germanic and Slavic languages (or opaque languages), for which null results are obtained with bare nouns or with any noun phrase not containing variable elements of gender agreement.

Although the GIE can be ground-breaking in the way we conceive the representation of grammatical gender, we have expressed some concerns about this effect in the previous chapter of this thesis. On the one hand, we have claimed that the effect was systematically obtained by the same research team, and thus replication with the same and other similar languages would be desirable. In 2011, Finocchiaro and colleagues tried to do so for Spanish, Italian, and French, but failed. On the other hand, the effect is itself problematic. Theoretically, it is expected that gender nodes compete with each other for selection, and hence, nouns of the same gender should facilitate gender retrieval in comparison to nouns of different gender, since the threshold for selection would be reached faster for the former. In an attempt to explain the unexpected GIE, Cubelli et al. (2005) proposed a theoretical approach, the DSM, based on the concept of phonological gender transparency. However, as we have seen, this proposal has raised criticisms. The notion of gender as a quasi-semantic feature at a common conceptual-grammatical lemma level seems incompatible with results from other types of studies on grammatical gender. Studies on gender priming hardly observe gender effects, and when they do, they are of facilitation and do not suggest the existence of any type of competition by similarity or the existence of a cohort of nouns of the same gender competing for selection (e.g., Heim, Eickhoff, et al., 2009; Vigliocco et al., 2002). Of particular concern is the fact that the GIE is in conflict with the cross-linguistic GCE in bilinguals, in which nouns of the same gender seem to facilitate lexical access rather than hindering it (e.g., Manolescu & Jarema, 2015; Morales et al., 2011; Paolieri, Cubelli, et al., 2010). Additionally, the model has been criticized because of the way on which authors equate phonological gender regularities ("-a" in "mesa") to gender morphemes; gender morphemes are in fact restricted to nouns with natural gender ("gat-o", "gat-a"). The idea that gender must be processed to encode the ending of a noun with abstract grammatical gender, such as "mesa", in which the "-a" is a mere thematic vowel, is controversial; yet more so if we consider that opaque nouns have also been shown to entail gender effects (Cubelli et al., 2005; Paolieri, Lotto, et al., 2010; Paolieri et al., 2011) and that these have final letters that do not correlate with any gender value (to better assess these criticisms, see Finocchiaro et al., 2011). Therefore, more research is necessary to understand to what extent their findings and explanations are generalizable, especially considering these and

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

other transparent languages. Our specific aim is to study grammatical gender processing within the PWI paradigm with another Romance language that is similar to Spanish and Italian in terms of transparency, European Portuguese.

Besides looking for the GIE in such a complex mixture of results, the examination of the impact of other variables on findings may shed further light on the debate. Indeed, we believe that one of the variables that may be relevant to debates here is animacy. Yet the animacy status of the experimental stimuli of all studies in this area has been systematically ignored, not only in the experimental designs, but also indirectly in terms of the experimental control of the materials. This is surprising, because: (1) the core of gender systems is closely linked to animacy (Corbett, 1991; Dahl, 2000); (2) it has been suggested that this link is especially relevant in languages with dual-gender systems (i.e., Romance languages; Sedlmeier et al., 2016); (3) our linguistic system seems to have a bias to assign animates to syntactically prominent positions, as previously noted (Branigan et al., 2008); and (4) the animacy of stimuli (words and images) has shown consistent cognitive repercussions in our perception of the surroundings and our responses to the world (New et al., 2007, 2010). Thus, bearing in mind that grammatical gender includes both animate and inanimate nouns, it seems crucial to consider animacy as a significant factor in the study of grammatical gender representation and processing. In what follows, besides looking for the GIE, we will address the possible cognitive and linguistic repercussions of including animate nouns in the design of an experiment along with inanimate ones. Several experiments in which the percentage of animate targets was manipulated (0%, 25%, 50%, and 100%) were conducted to examine the role of this variable in gender processing.

6.2.1 The role of animacy in the picture-word interference (PWI) paradigm and gender processing

When conducting a PWI task, the inclusion of animate nouns in the stimuli list might have significant implications on the outcome of the experiments, and even on the predictions of the models mentioned in the previous section. Research in cognitive psychology exploring human memory, perception, visual attention, and language processing has shown that our mind exhibits a clear preference for animate over inanimate stimuli (the so-called animacy effect, see Félix et al., 2019, for a recent overview). For instance: a) animate words are better recalled than inanimate words (e.g., Bonin et al., 2014; Félix et al., 2019; Leding, 2019; VanArsdall et al.,

USC UNIVERSIDADE DE SANTIAGO

2014); b) there is a bias for animate stimuli when deciding where to direct our visual attention, regardless of the degree of the perceived threat (Altman et al., 2016; Lipp et al., 2004; New et al., 2010); c) some grammatical effects are stronger when animate nouns are at issue (Dank et al., 2015); d) in lexical decision tasks, legal non-words perceived as referring to animate entities yield faster decision times than those perceived as inanimate (Bonin et al., 2019)[33]; and e) animate words seem to be acquired faster and more accurately than inanimate ones (Corrêa et al., 2011; Corrêa & Name, 2003). Animate stimuli, then, have a special place in our mind, being prioritized across all cognitive processes.

From a psycholinguistic point of view, animacy is considered to be one of the basic features of semantics, and has been attributed a special processing advantage, at least at the level of conceptual encoding during lexical access, as shown in many semantic categorization and violation studies (Bonin et al., 2019; Radanović et al., 2016; Xiao et al., 2016). In this sense, animate nouns tend to be semantically richer than inanimates, bearing more semantical features in their representations, and inducing a "deeper" conceptual processing of words, which consumes more resources and implies greater processing costs when semantic violations have to be detected (Szewczyk & Schriefers, 2011; Xiao et al., 2016). In terms of gender, it has been shown for language learners that gender agreement with animate nouns is cognitively more demanding than with inanimate nouns (Sagarra & Herschensohn, 2011). Thus, we might think that this deep semantic processing of animates during lexical access could have repercussions at other levels of word processing, namely the lemma, in which grammatical gender is selected. In this sense, animacy might be a reason for skipping gender selection during the processing of bare nouns. So, if animacy is a factor that encourages and enriches conceptual processing due to evolutionary pressures (Branigan et al., 2008; New et al., 2007), a direct connection between the conceptual and morpho-phonological strata, might be preferred, and hence gender processing would be skipped if agreement is not required. This is not an unheard idea, as was already proposed by Caramazza and Miozzo in their IN model (1997). According to this model, the features belonging to the grammatical-syntactic stratum are only activated/retrieved when required and so lexical access can occur through direct connections between the conceptual and the morpho-phonological levels without the grammatical-syntactic stratum intervening. In the

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

[33] But see Vigliocco and Frank (1999), and Vigliocco and Franck (2001) for evidence of no differences between animate and inanimate nouns in the percentage of agreement errors during language production.

case of animacy, the skipping of gender would allow the system to avoid unnecessary costs and prioritize the production of animate nouns.

From the perspective of other areas of cognitive psychology, the selective nature of animacy has been the basis of a fundamental theory of human attention supported by vast experimental evidence: The Animate Monitoring Hypothesis (New et al., 2007). Animates, it is claimed, possess an attentional advantage driven by ontogenetic factors that explains why participants are consistently faster and more accurate at detecting changes in animals (human and non-human) relative to changes in inanimate objects, even when these may constitute a threat for survival (e.g., vehicles). This attentional advantage entails a disadvantage for inanimate stimuli since their detection is distracted by the presence of animates (Altman et al., 2016). In fact, animate entities displayed in images have been shown to recruit visual attention in a way that is highly independent from the context.

In light of these findings from different areas of human cognition, it is not difficult to imagine the implications that animacy might have for a PWI task. On the one hand, the prioritization of conceptual encoding in animates may suppose the skipping of some grammatical encoding characteristics as structurally supported by the IN model. If grammatical gender processing is bypassed for either a target or a distractor noun due to their animacy status, it would not be possible to observe gender effects, since there would be no competition between gender nodes. On the other hand, following the attentional and perceptive repercussions of animacy, the degree of attention given to a target and a distractor may be highly disproportional if one of them is animate. In the case of animate targets, their supposedly high independence from the context may hamper the detection and processing of any distractor. Even if detected, the activation reached by the target noun may be notably higher than that reached by the distractor word, which might then be insufficient to generate any kind of competition. At the same time, animate distractors paired with inanimate targets may also be a source of disruption on the reliability of the results. This is so because any interfering effect may be substantially greater if the degree of activation reached by the distractor is "abnormally" high. However, if both targets and distractors are animate, gender effects might also be stronger, since competitive effects would imply higher levels of activation for both targets and distractors.

USC UNIVERSIDADE DE SANTIAGO

Despite this, on reviewing the literature on grammatical gender representation and processing during language production, it becomes clear that the role of animacy has been greatly neglected, which in turn might call the rationales developed over years regarding gender

effects into question. More specifically, it is a common practice in such studies to include random animate nouns in the stimuli list (targets and distractors) of a PWI paradigm in a non-proportional way across conditions. Only a few works have included zero animate nouns. In these, a GCE was observed in French with a positive Stimuli Onset Asynchrony (SOA) of 200 ms (Foucart et al., 2010) but not with a SOA of 0, either in French or in Hebrew (Alario & Caramazza, 2002; Dank & Deutsch, 2015). Yet the majority of studies on this matter include 8% to 33% of animate targets, and 3% to 24% of animate distractors (Costa, Kovacic, Fedorenko, et al., 2003; Cubelli et al., 2005; Finocchiaro, 2013; Finocchiaro et al., 2011; La Heij et al., 1998; Paolieri, Lotto, et al., 2010; Paolieri et al., 2011; O'Rourke, 2007; Schiller & Caramazza, 2003; Schriefers, 1993; Schriefers & Teruel, 2000; van Berkum, 1997; but no information is provided on animacy nor stimuli are available in the appendices of Costa et al., 1999; Miozzo & Caramazza, 1999; Miozzo et al., 2002). The percentages of animate stimuli are especially high in Cubelli et al. (2005), Paolieri, Lotto, et al. (2010), and Paolieri et al. (2011) studies, which encounter an unexpected GIE (up to 33.33% of targets and more than 20% of distractors; Cubelli et al., 2005; Paolieri, Lotto, et al., 2010; Paolieri et al., 2011).[34] Inconsistent results associated with these studies may be explained, at least partially, by this. If animacy does have a role in gender processing, future research should more effectively assess or control this variable and revise previous findings.

6.2.2 The present study

In this study, we aimed to explore the representation and processing of grammatical gender in a transparent Romance language that has received less attention, European Portuguese, while assessing the impact of animacy in gender retrieval within a PWI paradigm. To this end, we conducted two PWI experiments with native speakers of European Portuguese. In both experiments, bare nouns were used as a naming instruction. This allowed us: a) to test the GIE in another transparent Romance language; and (b) to analyse how animacy impacts the performance of participants as predicted by its unique nature in terms of semantics (Branigan et al., 2008) and the Animate Monitoring Hypothesis (New et al, 2007).

As one of our main purposes was to directly study grammatical gender selection during the production of animate nouns, for the sake of simplicity we focused on the use of animate targets,

[34] Note that in Paolieri, Lotto, et al. (2010), Paolieri et al. (2011), and Cubelli et al. (2005), experiments with 25% or more of animate targets (n = 6 experiments) have an average effect of 15.67 ms, whereas those with less than 25% of animate targets (n = 7 experiments) show an effect of 30.67 ms.

rather than distractors, which were kept inanimate. Therefore, we manipulated the percentage of animate targets included within each experimental list, emulating previous literature in which different amounts of animate nouns were included as stimuli (see Table 14 for the main conditions within each experiment). In Experiment 1, three different percentages of animate target nouns conveyed three different conditions: 0%, 50%, and 100%. Since European Portuguese is a Romance language with a high degree of phonological gender transparency, quite similar to Spanish, it fits the postulates of the DSM, by which nouns should retrieve their gender value even in the absence of an agreement context in order to process the nominal ending. Hence, following the tenets of the DSM, as well as the results obtained by Cubelli et al. (2010), Paolieri, Lotto, et al. (2010), and Paolieri et al. (2011), we expected to obtain competitive effects in the form of a GIE if the percentage of animate targets was low enough. As the percentage of animates varied across conditions, we expected the size of gender effects also to vary accordingly, as anticipated by either the semantic prioritization of animate nouns or the Animate Monitoring Hypothesis. Both approaches, although for different reasons, predict the same outcome: the higher the quantity of animate targets, the smaller the chance of finding competitive effects based on gender, or the smaller the effect sizes. In this sense, we expected no gender-based effects for the 100% animate condition, and indeed this was the case. Yet the condition of 50% did not show gender-based effects either. An analysis of the whole set of responses across the three conditions with the factor Target Animacy reveals that gender competitive effects were restricted to inanimate targets. In Experiment 2, we conducted another PWI task, this time with 25% of animate targets, to better understand when gender-based effects are obtainable within a PWI paradigm as a function of the percentage of animate targets within the list. The results failed to reach significance again. We will discuss two possible hypotheses that may be explaining our results and propose future research ideas to put them to the test. The data and scripts used in the study are available at the following link: https://osf.io/8px2z/.

Table 14

Conditions in each of the experiments

	Animacy presence	**Target gender**	**Gender congruency**	example (target-distractor)
Experiment 1	0% animate nouns	Masculine	Gender congruent	casaco-moinho [jacket-windmill]
			Gender incongruent	casaco-barriga [jacket-belly]
		Feminine	Gender congruent	mesa-praia

				[table-beach]
			Gender incongruent	mesa-palco [table-stage]
	50% animate nouns	Masculine	Gender congruent	queijo-sismo [cheese, earthquake]
			Gender incongruent	queijo-relva [cheese-grass]
		Feminine	Gender congruent	vaca-corda [cow-rope]
			Gender incongruent	vaca-figo [cow-fig]
	100% animate nouns	Masculine	Gender congruent	sapo-trigo [toad-wheat]
			Gender incongruent	sapo-rolha [toad-cork]
		Feminine	Gender congruent	cabra-túnica [goat-robe]
			Gender incongruent	cabra-astro [goat-star]
Experiment 2 (25% animate nouns)		Masculine	Gender congruent	dinossauro-telescópio [dinosaur-telescope]
			Gender incongruent	dinossauro-ferramenta [dinosaur-tool]
		Feminine	Gender congruent	bicicleta-camisola [bike-sweater]
			Gender incongruent	bicicleta-catálogo [bike-catalog]

6.2.3. Experiment 1

6.2.3.1 Method

6.2.3.1.1 Participants

Thirty-six native European Portuguese students from the University of Minho (31 female; M_{age} = 19.53 years, *SD* = 1.65) participated in the experiment and were rewarded with extra course credits. All signed an informed consent for experimentation with human subjects previously approved by the Ethics Council of the University of Minho (CEICSH 052/2019). None had moderate to high proficiency in another language with grammatical gender.

USC DE SANTIAGO DE COMPOSTELA

6.2.3.1.2 Materials

Pictures were selected from the International Picture Naming Project (IPNP) database (Szekely et al., 2004). We created a factor of Animacy Presence featuring three conditions. For the condition of 0% animates, 40 pictures representing inanimate objects were

selected. For the condition of 100% animates, another 40 pictures representing animate entities were selected. For the third condition of 50% animates, to maintain the control across conditions and types of pictures, 48 pictures from the previous two conditions (24 inanimates and 24 animates) were selected. Targets were distributed evenly across gender values within each of the three conditions, so that half of them were masculine and the other half feminine. In the condition of 50%, animacy was taken into consideration, and hence half of the animate nouns were masculine (e.g., "morcego" [bat]), the other half feminine (e.g., "raposa" [fox]) and the same applied to inanimate nouns (e.g., "sapato" [shoe] and "toalha" [towel]). This set of experiments does not have the aim to assess gender transparency and hence this variable was simply controlled as ortho-phonological gender transparency ("-a", /ɐ/ for the feminine; "-o", /u/ for the masculine). More specifically, in the 0% and 100% conditions, two feminine nouns were opaque (e.g., "avestruz" [ostrich]), and the other 18 were transparent for gender (e.g., "raposa"), whereas two masculine nouns were opaque (e.g., "peixe" [fish]), and the other 18 were transparent for gender (e.g., "morcego"). In the 50% condition, we maintained the same proportion of transparent and opaque nouns across the genders of the targets (22 transparent and 2 opaque per gender), but from the 4 opaque nouns, one of the opaque feminine nouns and one of the opaque masculine nouns were animate ("avestruz" [F] and "peixe" [M]), and the other two were inanimate ("chaminé" [F, chimney] and "tapete" [M, rug]). Nouns with natural (i.e., optional) gender were not used.

Pictures within the three conditions of animacy presence were matched for visual complexity (objectively defined by the digital size of the drawing) and goodness-of-depiction (i.e., how well each picture illustrated the target nouns), as obtained from the dataset in Szekely et al. (2005). Target words within each condition were matched across genders on multiple variables, as displayed in Table 15: per-million frequency, number of phonological and orthographic neighbours (N), word length (in letters), and mean logarithmic bigram frequency (these values were taken from the P-PAL database; Soares, Iriarte, et al., 2018); per-million frequency (Log10) values were also obtained from the SubtLex-PT database (Soares et al, 2015). Also, values of imageability and concreteness were obtained from the Minho Word Pool database (Soares et al., 2017). All *ps* > .115. Likewise, these variables were controlled: (1) in the 50% condition, taking into consideration the gender of the picture but also the animacy of the picture as an extra factor (*p* > .120), and (2) between the three conditions of animacy presence (all *ps* > .102). Importantly, differences between the conditions of animacy presence

USC UNIVERSIDADE DE SANTIAGO

on imageability and concreteness were analysed with 89% of the targets, as the remaining 11% were not present in the MWP database (see Table 2 for mean values of every controlled variable per target type). The analysis regarding the control of all the target pictures and nouns was conducted through a one-way analysis of variance (ANOVA) considering all the levels of the factors involved (within conditions of Animacy Presence: Target Gender and Gender Congruency, plus Animacy Status of the Target for the condition of 50% animates; and between conditions of Animacy Presence: Target Gender, Gender Congruency, and Animacy Presence).

Table 15

Mean and standard deviations (in parenthesis) of the controlled variables per target type for all the experimental pictures in Experiment 1

	Animacy 0%		Animacy 50%		Animacy 100%	
	Masculine	Feminine	Masculine	Feminine	Masculine	Feminine
FpM (log)	3.13 (0.54)	3.10 (0.54)	2.85 (2.35)	2.79 (0.39)	2.84 (0.72)	2.80 (0.63)
PN	4.20 (6.53)	5.15 (5.86)	2.75 (5.58)	4.67 (5.78)	4.05 (6.80)	3.30 (4.16)
ON	4.15 (5.59)	6.60 (7.42)	2.29 (4.16)	3.92 (5.24)	3.20 (4.83)	2.45 (3.40)
L	6.20 (1.70)	6.25 (1.52)	7.04 (1.96)	6.58 (1.52)	6.95 (1.92)	6.30 (1.39)
MLBF	2.57 (0.49)	2.53 (0.49)	2.39 (0.39)	2.36 (0.44)	2.50 (0.31)	2.31 (0.40)
VC	17,300.15 (9,922.55)	16,620.15 (6,975.19)	18,012.25 (8,685.89)	18,219.13 (8,023.40)	16,695.60 (5,653.05)	18,345.10 (7,573.65)
GoD	5.58 (0.76)	5.97 (0.63)	5.97 (0.54)	6.00 (0.64)	5.96 (0.63)	6.07 (0.45)
I	5.71 (1.03)	5.99 (0.40)	6.13 (0.29)	6.01 (0.34)	6.04 (0.36)	6.07 (0.36)
C	6.39 (0.89)	6.58 (0.18)	6.61 (0.20)	6.60 (0.20)	6.52 (0.22)	6.57 (0.18)

Note. FpM (log) = Frequency per million (Log_{10}). Obtained from the SubtLex-PT database (Soares et al., 2015). PN = Phonological number of neighbours; ON = Orthographic number of neighbours; L = Length (in letters); MLBF = Mean log bigram frequency. All obtained from the P-PAL database (Soares, Iriarte, et al., 2018). VC = Visual complexity; GoD = Goodness-of-depiction. All obtained from International Picture Naming Project (IPNP) database (Szekely et al. 2004). I = Imageability; C = Concreteness. Obtained from the Minho Word Pool database (Soares et al., 2017). The mean values for imageability and concreteness were calculated using the 82.5%, 87.5%, and 97.5% of target nouns of condition 0%, 50%, and 100% respectively, as the rest of nouns were not available in the database.

For each picture, four European Portuguese distractor nouns were selected (160 for each of the 0% and 100% conditions of animacy presence; 192 for the 50% condition). Two of these had the same grammatical gender as that of the picture (gender congruent condition), and the other two had different grammatical gender (gender incongruent condition). In addition, we controlled the transparency category of the distractors. Hence, from the four distractor nouns selected per target, one of the masculine distractors had the same transparency category as that of the target (e.g., both were transparent) and the other one had a different transparency category (e.g., target was transparent and distractor opaque or vice-versa). The same applied to feminine distractors. All distractors were inanimate nouns. Note that all target-distractor pairs were from different semantic categories, and there was no significant orthographic overlap between them across conditions, according to the NIM database (Guasch et al., 2013; all *ps* > .113 within conditions of animacy and *p* > .320 between conditions of animacy). Importantly, to avoid effects of facilitation during the naming task, all targets and their respective distractors differed in their initial phoneme (Kinoshita & Verdonschot, 2021; Mousikou et al., 2010; Schiller, 2004).

Targets were assigned to two experimental conditions as a function of their relation in gender congruency (GC) with the distractor. At the same time, a proportional distribution on transparency congruency was maintained for reasons of control. For instance, with the feminine transparent target "abelha" (bee), four distractors were selected: a) congruent in gender and in transparency category, such as "peruca" (wig); b) incongruent in gender but congruent in transparency, such as "estojo" (pencil case); c) congruent in gender but incongruent in transparency, such as "hélice" (propeller); and d) incongruent in both gender and transparency category, such as "quiosque" (kiosk). The same was applied to opaque feminine targets and to transparent and opaque masculine ones.

Distractors were matched within the three conditions of animacy presence through a one-way ANOVA conducted with the combination of each of the two levels of "target gender" and "gender congruency" on the number of phonological and orthographic neighbours (N), and word length (in letters), this obtained from the P-PAL database (Soares, Iriarte, et al., 2018), and on imageability and concreteness, this obtained from the MWP (Soares et al., 2017) (see mean values within each condition in Table 16, all *ps* > .08). Distractors were also matched between the three conditions of Animacy Presence in the same variables through a one-way ANOVA conducted with the combination of all the levels on the factors Target Gender, Gender

USC UNIVERSIDADE DE SANTIAGO

Congruency between target and distractor, and Animacy Presence" (all *ps* > .161). Importantly, the analysis revealed significant differences in the number of orthographic neighbours and in the number of letters (p = .017 and p = .018, respectively), yet Bonferroni post-hoc pair-wise comparisons actually revealed no significant differences between any condition (all *ps* > .07). Importantly, as discussed in the results, all variables plus logarithmic per million frequency as obtained by the SubtLex-PT database (Soares et al., 2015) were tested as fixed effects in the models of the analyses and none interacted with any of the effects.

Table 16

Mean and standard deviations (in parenthesis) of the controlled variables for the distractors across conditions of Target Gender and Gender Congruency within every condition of Animacy Presence in Experiment 1

	Animacy 0%				Animacy 50%				Animacy 100%			
	MGC	MGI	FGC	FGI	MGC	MGI	FGC	FGI	MGC	MGI	FGC	FGI
FpM (log)	2.80	2.70	2.77	2.63	2.56	2.44	2.46	2.34	2.52	2.26	2.34	2.37
	(.72)	(.68)	(.58)	(.68)	(.58)	(.70)	(.59)	(.56)	(.75)	(.84)	(.86)	(.77)
PN	3.50	3.43	4.68	4.30	2.21	2.31	3.73	3.35	3.15	3.55	2.65	2.68
	(4.30)	(4.49)	(4.93)	(4.24)	(2.91)	(3.48)	(4.11)	(3.79)	(4.43)	(5.00)	(2.59)	(3.54)
ON	3.13	3.40	4.53	4.43	1.75	2.04	3.52	3.02	2.53	3.05	2.40	2.20
	(4.05)	(4.11)	(4.87)	(4.81)	(2.40)	(2.97)	(4.30)	(3.92)	(3.42)	(4.24)	(3.04)	(3.04)
L	6.20	6.15	6.13	5.95	7.08	6.92	6.56	6.56	6.93	6.90	6.53	6.40
	(1.71)	(1.64)	(1.64)	(1.52)	(1.71)	(1.83)	(1.39)	(1.39)	(1.77)	(11.89)	(1.33)	(1.33)
I	5.62	5.40	5.40	5.47	5.68	5.34	5.63	5.56	5.69	5.51	5.36	5.59
	(0.70)	(1.20)	(1.08)	(0.89)	(0.82)	(1.01)	(0.82)	(0.82)	(1.08)	(1.20)	(1.45)	(0.89)
C	5.89	5.56	5.52	5.92	6.03	5.68	5.96	5.91	6.00	5.85	5.73	5.88
	(1.08)	(1.77)	(1.58)	(0.89)	(1.08)	(1.64)	(1.20)	(1.14)	(1.14)	(1.71)	(1.90)	(1.20)
OV	0.15	0.12	0.14	0.15	0.16	0.13	0.14	0.13	0.14	0.14	0.14	0.09
	(0.06)	(0.13)	(0.06)	(0.13)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.13)	(0.13)	(0.06)

Note. M and F = Masculine and Feminine. GC and GI = Gender Congruent and Gender Incongruent. FpM (log) = Frequency per million (Log_{10}). Obtained from the SubtLex-PT database (Soares et al., 2015). PN = Phonological number of neighbours; ON = Orthographic number of neighbours; L = Length (in letters). All obtained from the P-PAL database (Soares, Iriarte, et al., 2018). I = Imageability; C = Concreteness. Obtained from the Minho Word Pool database (Soares et al., 2017). OV = Orthographic overlap between targets and distractors. Obtained from the NIM database (Guasch et al., 2013). The mean values for imageability and concreteness were calculated using the 66%, 54%, and 51% of distractor nouns of condition

0%, 50%, and 100% respectively, as the rest of nouns were not available in the database. Note that the values for FpM (log) showed significant differences in the ad hoc ANOVA but the variable was included in the final models and did not hold any type of impact on the results.

Finally, as training items, eight different pictures were selected from the IPNP database, along with eight new distractors. Within each condition of Animacy Presence, four different lists were created for counterbalancing purposes. Each of these featured the same conditions based on the Target Gender (masculine/feminine) and the Gender Congruency between target and distractor (congruent/incongruent), including the 40 (0% and 100% of Animacy Presence) or 48 target pictures (50% of Animacy Presence). Each target picture was presented four times within each condition of Animacy Presence, one time per list and each time associated with a different distractor. Participants were randomly assigned to one of the four lists, assuring the same number of participants per list.

6.2.3.1.3 Procedure

Participants engaged individually in the three different conditions of animacy presence (0%, 50% and 100%, of animate targets), i.e., all did the three conditions, these separated by a month and a half time lapse. We guaranteed that the first condition for all participants was the one containing no animate target, since we wanted to avoid any kind of confounding effects of animacy emerging from previous exposure to the task with pictures showing animate entities. Given that the condition of 50% animate targets contained the stimuli from the other two conditions, we also considered that this should be the last one to be conducted, so that animate and inanimate targets were not disproportional in their novelty for the participants. After this, a familiarization phase and an experimental task took place in a soundproof booth, and these were subsequently repeated for each condition, but with different materials. Participants signed a consent form for the whole experiment before the first session.

The familiarization phase consisted of the presentation of all the pictures involved in the subsequent experimental task to guarantee high naming agreement scores. Each picture was presented automatically for 2 s in a single slide, along with the nouns that participants would have to use to name them in the task. Words were shown in lowercase letters (Agency FB 36) below the respective picture. The presentation was done in Microsoft Powerpoint (2010). No overt response was required. Stimulus presentation was randomized across participants.

USC UNIVERSIDADE DE SANTIAGO

The experimental task consisted of a PWI paradigm, in which participants were instructed to name each picture using a bare noun (“mesa”) as quickly and accurately as possible, while

ignoring the distractor noun superimposed onto the picture. All the pictures were presented at the centre of the computer screen. The distractors, presented at the centre of the pictures with a 0 ms SOA, were shown in lowercase letters in Courier New font, 14 point (see Cubelli et al., 2005 for a similar procedure). Experimental trials followed this structure: (a) a fixation point (+) at the centre of the computer screen, for 500 ms; (b) the target picture with the superimposed distractor for 2,000 ms or until response; (c) a blank space for 500 ms as an inter-trial interval. Participants first performed a training block of 8 trials. Experimental trials were presented pseudo-randomly per participant following the next rules: only 2 targets of the same Gender Congruency condition, same Target Gender condition, and the same controlled Transparency Congruency variable (transparent/congruent targets and distractors) could appear consecutively. Response times (RTs) were measured from the onset of the stimulus to the beginning of the naming response. Stimulus presentation was done using the DMDX software (Forster & Forster, 2003). Naming RTs were recorded by the voice-key from the presentation of the target to the onset of the naming response and then checked offline using the CheckVocal software (Protopapas, 2007).

The session for each condition of animacy lasted approximately 10 min.

6.2.3.2 Results

We discarded 491 data points (10.66% of the total) from the analyses: those containing incorrect responses (357 data points, 7.75% of the total), and those with RTs that exceeded 2.5 *SD* of each participant's mean (134 data points, 2.91% of the total). Data were analysed by means of linear mixed-effect models (e.g., Baayen, 2008; Baayen et al., 2008). We used the lme4 package from R (Bates et al., 2014). Different linear models were created to examine the effect of the variables of interest (Animacy Presence, Target Gender and Gender Congruency) and their interactions on inverse RTs (-1000/RT). These variables and their interactions were introduced as fixed effects. In addition, each model included random intercepts and random slopes by participants and items, following a maximal random effects structure (Barr et al., 2013). In most cases, we were unable to get the maximal random effects structure to converge, so we kept only those random slopes that allowed for convergence. In particular, the random structure was stepwise simplified by removing in each step the effect that explained the smallest proportion of the variance, starting from the most complex model (i.e., that including the three-way interaction in Experiment 1, and the two-way interaction in Experiment 2), until a

USC UNIVERSIDADE DE SANTIAGO

convergent model with at least one critical effect on the random slopes was obtained. To determine the significance of fixed effects, log-likelihood ratio tests were used (R function Anova). We assessed the contribution of each fixed effect by comparing a model that included the effect of interest with another model that did not include it. P-values for pairwise comparisons were estimated using the7Test package (Kuznetsova et al., 2014).[35]

Table 17

Mean RTs and standard errors (SE) for each condition in Experiment 1

Animacy	**Gender**	**Congruency**	**Mean**	***SE***	***GCE size***
0%	Masculine	Incongruent	879	12.00	
		Congruent	856	11.10	
	Feminine	Incongruent	880	11.50	
		Congruent	855	11.60	
					24 ms
100%	Masculine	Incongruent	896	11.60	
		Congruent	909	12.30	
	Feminine	Incongruent	918	12.50	
		Congruent	917	12.00	
					-6 ms
50%	Masculine	Incongruent	829	10.40	
		Congruent	815	8.87	
	Feminine	Incongruent	837	9.96	
		Congruent	851	10.80	
					0 ms

Note. Animacy stands for factor "Animacy Presence", Gender stands for factor "Target Gender", Congruency stands for factor "Gender Congruency". Formula of the final model: -1000/RT ~ congruency*animacy*gender + (animacy | participant) + (congruency | item).

Table 17 shows the mean RTs for each condition. The interaction between Gender Congruency and Animacy Presence reached significance, $\chi^2(2) = 7.51$, $p = .023$. To examine this interaction, we analysed the GCE in each of the conditions of Animacy Presence. There was a GCE in the 0% animates condition, *Estimate* = -0.029, *SE* = 0.011, $p = .008$, $\chi^2(2) = 7.15$, $p = .007$, but no effects of congruency were observed in the other two animacy conditions (all *p*s > .292). In addition, we found a significant effect of animacy, $\chi^2(2) = 26.45$, $p < .001$,

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

[35] In each of the three conditions, as well as in the next experiment, we also created a series of models in which we introduced as fixed effects the variables that were controlled between the different conditions (e.g., word frequency, number of neighbors, etc.; see Materials). In many cases, these models did not converge, and in those cases in which they did, the results did not differ from the models that did not include the controlled variables. In view of this, and for the sake of clarity and simplicity, we decided to present the results of the models that did not include the controlled variables.

indicating that RTs for the 50% animates condition were faster than both RTs for 0% animates condition, *Estimate* = -0.082, *SE* = 0.016, $p < .001$, and RTs for the 100% animates condition, *Estimate* = -0.087, *SE* = 0.018, $p < .001$, probably due to an effect of familiarization with the viewed images (note that the 50% condition was composed of stimuli taken from the other two conditions and was the last carried out by the participants). No other main effects or interactions were observed.

A possible explanation for the gender effect found in the 0% animate condition, but not in the 50% and 100% animate ones, might be due to the fact that animate and inanimate targets were not perfectly matched in their lexical and semantic properties. We tested this hypothesis by examining the interaction between Gender Congruency and Target Animacy (animate vs. inanimate targets rather than the percentage of animates in the stimuli list) on the whole set of responses from the three conditions after controlling for all these variables. We created a linear mixed-effects model in which we introduced all lexical (e.g., frequency, number of neighbours, etc.) and semantic (e.g., concreteness) variables of the target words and pictures (e.g., visual complexity, goodness-of-depiction), together with the interaction between Gender Congruency and Target Animacy, as predictors of RTs. We also introduced participants and items as random effects, and gender congruency in the item random slope (any more complex structure of random effects did not allow for model convergence). The results showed an interaction between Gender Congruency and Target Animacy, $\chi^2(1) = 4.92$, $p = .027$. A significant congruency effect was observed for inanimate targets, *Estimate* = -0.09, *SE* = 0.03, $p = .003$, $\chi^2(1) = 5.10$, $p = .024$, but this effect was not significant for animate targets, $\chi^2(1) = 0.82$, $p = .366$. This has two strong implications: (1) It confirms that indeed, animate and inanimate targets may behave differently and that inanimate targets show gender competitive effects whereas animate ones do not; (2) it suggests that the results of this experiment do not seem to be due to the effect of the variables in which the animate and inanimate targets (and distractors) vary.

The GCE found in the 0% animates list is indeed surprising from the point of view of the previous empirical evidence and the DSM model. On the one hand, it is in line with the idea that in Romance/transparent languages gender is retrieved in the absence of agreement and gender nodes compete for selection, as supported by the GSIH. However, we not only failed to replicate the GIE but observed gender effects in the opposite direction. This direction, however, is expected given the conception of competition between gender nodes that WEAVER++

USC UNIVERSIDADE DE SANTIAGO

proposes and the results obtained in gender priming and bilingual studies (the cross-linguistic GCE). Hence, the results are favourable to the idea that gender is represented through two gender nodes that accumulate activation and compete for selection in a grammatical stratum of lexical access. On the other hand, as predicted by the semantic prioritization hypothesis, by which animacy could be responsible for skipping gender selection, and also by the Animate Monitoring Hypothesis, by which the degree of attention given to animate pictures could diminish the competitive role of the distractors, there was no competition between gender values of target and distractors for animate targets. Surprisingly, there was not even an effect for the list with 50% of animate and inanimate targets. In order to test for the minimum allowable percentage of animates within the stimuli list for gender effects to be obtained, we conducted a second experiment involving 25% of animate targets, this to include a greater range of cases and to understand the extent to which prior studies might have been affected by the random inclusion of animate targets.

6.2.4 Experiment 2

6.2.4.1 Method

6.2.4.1.1 Participants

Forty-eight native European Portuguese students from the same population as that of Experiment 1 (46 female; M_{age} = 21.32 years, *SD* = 3.88) participated in this experiment and were rewarded with extra course credits. All signed an informed consent for experimentation with human subjects previously approved by the Ethics Council of the University of Minho (CEICSH 052/2019). None had moderate to high proficiency in another language with grammatical gender.

6.2.4.1.2 Materials

We selected 48 pictures from Experiment 1: twenty-five percent of them (12) were animate and the other seventy-five percent (36) were inanimate. We selected four distractors per picture (a total of 192 distractors), following the same assignment rules previously described and keeping the proportion between transparent and opaque nouns. Pictures were controlled across gender values (masculine and feminine) for complexity and goodness of depiction (*p*s > .126); when considering the factor Animacy (animates and inanimates) along with Target Gender, pictures also showed no significant differences in those variables (*p*s > .434). As in Experiment 1, target nouns were matched across gender values on the variables: per-million

USC UNIVERSIDADE DE SANTIAGO

frequency (Log10), this obtained from the SubtLex-PT database (Soares et al., 2015); per-million frequency, number of phonological and orthographic neighbours (N), word length (in letters), and mean logarithmic bigram frequency, these obtained from the P-PAL database (Soares, Iriarte, et al., 2018); imageability and concreteness, these obtained from the Minho Word Pool database (MWP; Soares et al., 2017). All *p*s > .096 (see Table 18 for mean values). Note that the imageability and concreteness variables were only controlled for 85% of the target nouns, for which values were found in the Minho Word Pool database. Of interest, targets were also controlled when considered not only across gender values but also across animates and inanimates (all *p*s > .267). All these analyses were based on one-way ANOVAS.

Table 18

Mean and standard deviations (in parenthesis) of the controlled variables across targets' gender values for all the experimental pictures in Experiment 2

	Masculine	**Feminine**
FpM (log)	3.02 (0.49)	2.95 (0.49)
PN	3.42 (5.98)	5.46 (6.07)
ON	3.42 (5.29)	5.67 (7.15)
L	6.58 (2.01)	6.38 (1.57)
MLBF	2.45 (0.44)	2.45 (0.49)
VC	16,067.79 (5,355.32)	18,857.88 (7,939.83)
GoD	5.95 (0.59)	6.08 (0.44)
I	5.88 (0.93)	6.00 (0.34)
C	6.44 (0.78)	6.58 (0.18)

Note. FpM (log) = Frequency per million (Log_{10}). Obtained from the SubtLex-PT database (Soares et al., 2015). PN = Phonological number of neighbours; ON = Orthographic number of neighbours; L = Length (in letters); MLBF = Mean log bigram frequency. All obtained from the P-PAL database (Soares, Iriarte, et al., 2018). VC = Visual complexity; GoD = Goodness-of-depiction. All obtained from International Picture Naming Project (IPNP) database (Szekely et al. 2004). I = Imageability; C = Concreteness. Obtained from the Minho Word Pool database (Soares et al., 2017). The mean values for imageability and concreteness were calculated using 85% of the target nouns, as the rest of nouns were not available in the database.

USC UNIVERSIDADE DE SANTIAGO

Distractors were controlled across conditions of Gender Congruency and Target Gender for the variables number of phonological and orthographic neighbours (N), and word length (in letters), these obtained from the P-PAL database (Soares, Iriarte, et al., 2018), and on imageability and concreteness (115, i.e., 60% of the distractors), these obtained from the MWP (Soares et al., 2017) (see mean values within each condition in Table 19, all *p*s > .152). Besides, all target-distractor pairs differed in their initial phoneme, were from different semantic categories, and there was no significant orthographic overlap between them across conditions, according to the NIM database (Guasch et al., 2013; $p > .710$). All these analyses were based on one-way ANOVAS. Importantly, as said in the previous experiment, all variables plus logarithmic per million frequency as obtained by the SubtLex-PT database (Soares et al., 2015) were tested as fixed effects in the models of the analyses and none interacted with any of the effects.

Table 19

Mean and standard deviations (in parenthesis) of the controlled variables for the distractors across conditions of Target Gender and Gender Congruency in Experiment 2

	MGC	MGI	FGC	FGI
FpM (log)	2.76 (.46)	2.54 (.69)	2.64 (.55)	2.51 (.61)
PN	3.04 (4.02)	3.08 (4.16)	4.44 (4.71)	4.35 (4.43)
ON	2.63 (3.60)	3.04 (3.46)	4.10 (5.06)	4.02 (4.71)
L	6.48 (1.73)	6.56 (1.94)	6.42 (1.59)	6.10 (1.52)
I	5.65 (0.83)	5.28 (1.18)	5.52 (0.97)	5.45 (0.90)
C	6.01 (1.04)	5.43 (1.80)	5.77 (1.45)	5.97 (0.97)
OV	0.14 (0.07)	0.17 (0.35)	0.13 (0.07)	0.12 (0.07)

Note. M and F stand for Masculine and Feminine. GC and GI stand for Gender Congruent and Gender Incongruent. FpM (log) = Frequency per million (Log_{10}). Obtained from the SubtLex-PT database (Soares et al., 2015). PN = Phonological number of neighbours; ON = Orthographic number of neighbours; L = Length (in letters). All obtained from the P-PAL database (Soares, Iriarte, et al., 2018). I = Imageability; C = Concreteness. Obtained from the Minho Word Pool database (Soares et al., 2017). OV = Orthographic overlap between targets and distractors. Obtained from the NIM database (Guasch et al., 2013). The mean values for imageability and concreteness were calculated using 60% of the distractor nouns, as the rest of nouns were not available in the database. Note that the values for FpM (log) showed significant

USC UNIVERSIDADE DE SANTIAGO

differences in the ad hoc ANOVA but the variable was included in the final models and did not hold any type of impact on the results.

6.2.4.1.3 Procedure

We followed the same procedure as that of Experiment 1, except that it was only one PWI session as there was no factor of Animacy Presence.

6.2.4.2 Results

We excluded 254 data points (11.03% of the total) from the analyses: those corresponding to incorrect responses (178 data points, 7.73% of the total), and those containing RTs that exceeded 2.5 *SD* of each participant's mean (76 data points, 3.30% of the total). The data analyses were identical to those performed in Experiment 1.

Table 20

Mean RTs and standard errors (SE) for each condition in Experiment 2

Gender	**Congruency**	**Mean**	***SE***	***GCE size***
Masculine	Incongruent	848	8.76	
	Congruent	862	9.41	
Feminine	Incongruent	847	8.81	
	Congruent	836	8.66	
				-2 ms

Note. Gender stands for factor "Target Gender", Congruency stands for factor "Gender Congruency". Formula of the final model: -1000/RT ~ congruency*gender + animacy + (congruency | participant) + (congruency | item).

The mean RTs for each condition are presented in Table 20. No main effect or interaction reached significance. This came out as surprising: with only 25% of animate targets within the stimuli list, the GCE disappeared, even though gender competitive effects have been observed with similar percentages of animate targets (e.g., Cubelli et al., 2005; Paolieri, Lotto, et al., 2010; Paolieri et al., 2011). The fact that such a small percentage of animate targets may be behind this absence gives animacy a greater impact than we could have imagined when inspecting both the theoretical and the experimental literature. To be sure of the null results, we examined the data through Bayesian analyses. We used the Bayes Factor (BF_{10}), which allowed us to quantify the amount of evidence for (H_1) and against (H_0) the effect of the variables of interest. The magnitude of this evidence is presented as an odds ratio (H_1 evidence/ H_0 evidence), which can range from 0 to infinite. If the value increases, it provides evidence in

favour of H_1; if it approaches 0, it provides evidence in favour of H_0. Values close to or equal to 1 indicate that both H_1 and H_0 are equally probable. By convention, values above 3 can be interpreted as moderate evidence supporting the H_1, and values below 1/3 give moderate support for the H_0 (Dienes, 2014; Jeffreys, 1961). We used the BayesFactor package of R (Morey & Rouder, 2015) to perform these analyses. We compared a model that included the factor of interest (H_1) with one that did not (H_0) using the lmBF function. Here we will use BF01 to indicate the amount of evidence in favour of H0 relative to H1 (in which the interpretation is inverted, i.e., BF01 of 3 would suggest moderate evidence for H0 and 1/3 moderate evidence for H1, and so on.). A JZS prior with scaling factor r = .707 was used in all analyses (Rouder et al., 2009).[36] The effect of gender congruency, $\chi^2(1) = 0.04$, $p = .832$, $BF_{01} = 10.86$, gender, $\chi^2(1) = 0.64$, $p = .425$, $BF_{01} = 4.04$, and the interaction between Gender Congruency and Target Gender, $\chi^2(1) = 3.04$, $p = .081$, $BF_{01} = 4.15$, were not significant and the Bayesian analyses suggest that the evidence for the null hypothesis is large enough, supporting the absence of effects in this experiment. Although a direct comparison between animate and inanimate targets as the one conducted with all the lists in Experiment 1 was not possible due to the low number of animate targets, we additionally examined the effect of gender congruency exclusively for inanimate targets. As in the analysis with all targets, the GCE was not significant, $\chi^2(1) = 0.22$, $p = .637$, $BF_{01}= 15.66$. Overall, the results contrast with those from other studies testing Romance languages in which a GIE was obtained (Cubelli et al., 2005; Paolieri, Lotto, et al., 2010; Paolieri et al., 2011) and the percentage of animate stimuli, and particularly targets, was similar to the percentage featured in this experiment (10% to 33.33% of animate targets plus animate distractors with unknown repercussions for gender processing). Two different but perhaps complementary underlying reasons for our results can be advanced here. On the one hand, we have seen that gender-based effects are quite small and slippery, not only in monolingual populations when assessing the GIE, but also in bilingual populations in which they are more consistently obtained (cross-linguistic GCE: $g = 0.24$, $SE = 0.04$, $p < .001$; native GIE: $g = -0.171$, $SE = 0.062$, $p = .021$). The absence of an effect with the inanimate targets of our additional analysis may be an indication of the lack of robustness of the effect. Also, as seen, the PWI paradigm is not ideal in terms of sensitiveness to detect certain effects of competition. Hence, gender effects such as our GCE may be highly sensitive to the list composition, and the 25% of animate targets here could indeed have diminished our chances

[36] We thank an anonymous reviewer for suggesting the use of Bayesian Analyses.

of observing any effect in this experiment. In this sense, a second possible cause behind the present results may be a carry-over effect within the list induced by animate targets. We will consider this in more detail in the Discussion.

6.2.5 Discussion

In the present study, we aimed to examine the gender (in)congruency effect in its interaction with the animacy of the stimuli, in an attempt to better understand the representation and processing of grammatical gender during language production. To that end, two PWI experiments with European Portuguese bare nouns were conducted. On the one hand, these experiments would allow us to explore the gender-based effects of competition predicted by the tenets of DSM model, according to which the production of bare nouns in a gender transparent language unavoidably entails the selection of gender. On the other hand, they would also allow us to test the impact of animacy on the predicted competitive effects, following the tenets of the semantic prioritization that occurs in the lexical access of animate nouns (Branigan et al., 2008) and those of the Animate Monitoring Hypothesis (New et al., 2007). We expected to obtain the GIE that was previously observed in other transparent Romance languages (Italian: Cubelli et al., 2005; Italian and Spanish: Paolieri, Lotto, et al., 2010; Paolieri et al., 2011). However, the GIE should be observed only if the percentage of animate targets was low enough, as animacy may induce the skipping of gender selection or/and capture the attention of participants, thus diminishing activation and the competitive role of the distractors.

In the first experiment, different percentages of animate target nouns were tested, with all distractors inanimate. The condition of 0% animate targets was critical, as it provided the ideal conditions to observe competitive gender effects. Indeed, competitive gender effects were obtained, although in the form of congruency rather than incongruency. This supports the original idea of models such as WEAVER++ and the IN model, by which grammatical gender is represented by gender nodes that accumulate activation and compete for selection. These competitive effects also suggest that the target noun is not only activating a gender node during lexical access but is also selecting it, which triggers competition for selection with the gender node activated by the distractor. Crucially, this happened without the presence of an agreement context, contrarily to the premises of WEAVER++ and what is observed with Germanic and Slavic languages (e.g., Schiller & Caramazza, 2003; Schriefers, 1993). The results, hence, constitute supporting evidence for the idea that transparent languages activate and select gender

USC UNIVERSIDADE DE SANTIAGO

at the level of grammatical encoding during bare noun production (GSIH). However, they go against the GIE predicted by the DSM (Cubelli et al., 2005), which was previously obtained in transparent languages that are similar to European Portuguese, such as Spanish and Italian (Cubelli et al., 2005; Paolieri, Lotto, et al., 2010; Paolieri et al., 2011). The DSM describes grammatical gender as an inherent characteristic of nouns located at a lemma level that integrates both conceptual and grammatical features. Every word would have its own lemma encompassing all its semantic and grammatical features; hence, word lemmas would include their specific gender value as an associated feature and, since they compete for selection by similarity, lemmas of the same gender interfere with each other during lexical access. Our results, though, suggest that gender values are represented by nodes (the masculine and feminine gender nodes) located at a lemma level or a syntactic network separated from the level of conceptual encoding. These nodes seem to gather activation until reaching the threshold for selection. The activation of different gender values, and thus different gender nodes, would translate into interference for gender selection. Similarly, nouns of the same gender would contribute to the activation of the same gender value, facilitating its selection. Our results also coincide with the architecture underlying lexical access as required by the cross-linguistic GCE obtained in bilingual populations, for whom competition does not occur by similarity (Klassen, 2016; Manolescu & Jarema, 2015; Morales et al., 2011; Paolieri, Cubelli, et al., 2010; Paolieri et al., 2019).

Regarding animacy, results were quite clear: gender competitive effects were restricted to inanimate targets and to the 0% animacy condition. The absence of effects in both the 50% animate condition and in Experiment 2 with 25% of animate targets comes out as striking. Indeed, as seen in the meta-analyses conducted in the previous chapter, competitive effects based on gender have been shown to be a small and heterogeneous both in monolingual and bilingual populations. This means that gender effects can be highly sensitive to the composition of the stimuli list. The analysis of the inanimate targets in Experiment 2 suggests that indeed, the GCE we obtained is not robust and that the presence of animate stimuli could hence have decreased our chance to observe it.

USC UNIVERSIDADE DE SANTIAGO

We advance two possible reasons underlying the absence of gender effects with animate targets that might also explain why effects disappear with the inclusion of animate targets in the stimuli list. On the one hand, following the Animate Monitoring Hypothesis and its supporting evidence (e.g., Altman et al., 2016; New et al., 2007), animates (especially if

represented by pictures) entail an attentional bias and are highly independent of the context. Therefore, the context remains easily ignored by our visual perceptive mechanisms if an animate entity is included. This could result in an attentional deficit for the distractors, causing an absence of effects with animate targets. On the other hand, the explanation could be on the character of animacy from a semantic perspective. More specifically, due to the high amount of resources that are consumed in the first phase of conceptual encoding and the "*special*" status that is given to animates for ontogenetic reasons, if agreement is not necessary, gender selection may be skipped at the lemma level to avoid unnecessary costs as well as avoiding slowing down the response. The idea of skipping the processing of grammatical characteristics when not necessary was already proposed in the IN model (Caramazza, 1997), which states that a response can be given through direct connections from the conceptual to the morpho-phonological stratum. These two hypotheses could explain the absence of gender competitive effects with animate nouns. Similarly, the inclusion of animate targets in the stimuli list could also be overshadowing the effect of the distractors beyond expected. Following the Animate Monitoring Hypothesis (New et al., 2007), if animate pictures are capturing the participants' attention in an accentuated way, the degree of attention given to inanimate pictures can also be higher, diminishing the interfering role of the distractors. The same could be said about the semantic prioritization principle of the processing of animate nouns. It may be the case that when gender processing is skipped for some stimuli, it induces the skipping of this characteristic for other stimuli within the same task. Although alterations in the processing of inanimate stimuli seem to better fit the hypothesis based on attentional mechanisms, it is worth mentioning that, when looking at the means for the animacy presence conditions (see Table 16), our RTs suggest that animates indeed require more cognitive resources and hence entail deeper processing, which would suggest prioritization given the rich semantic encoding. Specifically, the mean RTs appear to be higher for the condition of 100% than for the other conditions. Interestingly, in Experiment 1 the condition of 50% animates shows the lowest RTs and these are significantly different from the mean RTs in the conditions featuring 0% and 100% of animates. At first glance, we would expect the condition of 50% to show higher RTs than the condition of 0%, but lower than the condition of 100%. As mentioned above, this result can be explained simply as the effect of practice. The condition of 50% animates was conducted after the conditions of 0% and 100% animates and consisted of half of the stimuli of the other two conditions. Even though each session took place one month and a half after the previous one,

USC UNIVERSIDADE DE SANTIAGO

we believe that this is the most reasonable explanation in both theoretical and experimental terms. If we are right, both animate and inanimate stimuli should have benefited from the earlier presentation of the stimuli in the 0% and 100% animate condition, respectively. We compared the mean RTs of the inanimate stimuli from the 0% animate and the 50% and conditions, as well as those of animate stimuli from the 100% and 50% animate conditions. The results showed significant differences for both comparisons, in line with the explanation based on an effect of practice (868 vs. 819 ms, 910 vs. 848 ms, respectively, all *ps* < .001, see Table 21).

Table 21

Mean RTs and standard errors for animate and inanimate nouns in Experiment 1

Animacy Presence	**Animacy status**	**Mean**	***SE***
0% animates	Inanimates	868	5.78
100% animates	Animates	910	6.04
50% animates	Inanimates	819	6.63
	Animates	848	7.54

Note. Animacy status is related to the animacy status of the target nouns of each condition of Animacy Presence.

Importantly, the two hypotheses that we have presented as possible explanations to the absence of gender competitive effects in animate target nouns are not necessarily antagonistic. In this sense, an attentional bias for animates may exist, precluding the observation of any competitive effect; nevertheless, animate nouns may also skip gender processing due to semantic prioritization. Future studies should use other paradigms and experimental techniques (e.g., eye-tracking, event-related potentials) towards disentangling the two alternatives. More specifically, we encourage future studies to examine other type of effects, especially an ortho-phonological facilitation effect with animate targets and inanimate distractors. For instance, in an experiment in Portuguese, one could select animate targets such as "macaco" (monkey) and pair them with form-related distractors such as "casaco" (jacket) and unrelated distractors such as "*panela*" (pan). If attentional-driven reasons are behind the absence of results with animate targets in such a study, an ortho-phonological based effect should not be obtained or should at least decrease its size as the quantity of animate targets increases. On the contrary, if semantic-driven reasons are in the roots of our results and the processing of grammar features such as gender is skipped, form effects should emerge regardless of the animacy status of the targets. Additionally, the role of animacy in gender-based effects and in the PWI itself could also be explored in other languages, to understand to what extent animacy is relevant for gender

processing in languages in which what is obtained is a determiner-congruency effect that is also clearly based on gender.

Finally, the most problematic situation here arises directly from the type of gender competitive effects we have obtained, which are of congruency instead of incongruency, as observed by Cubelli et al. (2005), Paolieri, Lotto, et al. (2010), and Paolieri et al. (2011). We must consider two possible views for this disparity, depending on whether we question the reliability of the GIE or of the GCE. On the one hand, the GIE (1) is not only theoretically but also empirically difficult to explain if we take into consideration previous evidence on both gender priming and bilingual gender processing (i.e., the cross-linguistic GCE); (2) was obtained by the same research team and could not be replicated by another research team (Finocchiaro et al., 2011). We should hence ask ourselves what the differences between the studies finding the GIE and ours really are. An inspection of the materials in those studies confirms that the percentage of animate nouns in the targets list ranges from 10% to 33.33%, and that of the distractors between 5% and 25%, these having been randomly included. This leads us to question the extent to which these results may be an entirely reliable reflection of gender processing, especially if only inanimate nouns are responsible for the effect. It might be worth mentioning that there may be some differences in terms of stimuli control between these studies and our own. In this respect, our study had stricter control across conditions, featuring 21 variables related to both targets and distractors. Our experiments were also controlled for gender transparency of targets and distractors, and gender transparency congruency, semantic relatedness, and first phoneme overlap between targets and distractors. This makes leads us to wonder which variable other than animacy itself could be behind our results, especially taking into consideration that gender-based effects are indeed small and heterogeneous effects that are highly sensitive to the composition list. Perhaps the studies by Cubelli et al. (2005), Paolieri, Lotto, et al. (2010), and Paolieri et al. (2011) could indeed have lacked the necessary control to obtain reliable results (only 4 or 5 variables were controlled ad-hoc [word frequency, length, gender transparency, phonological overlap, and sometimes semantic relatedness], with no testing for inclusion in the models, and with no control between animate and inanimate stimuli). We believe that future research in Italian and Spanish should increase the control of the materials across conditions and feature only inanimate pictures and distractors in the experimental list. Likewise, the specific role of animate distractors in the competitive gender effects should be carefully assessed. Importantly, we should note that the first experiment of

USC UNIVERSIDADE DE SANTIAGO

Finocchiaro et al. (2011) used exactly the same materials as Cubelli et al. (2005) and failed to obtain the effects (the second and third experiments with Italian and Spanish participants used another set of stimuli and thus were not an exact replication). In any case, we cannot disregard the possibility that these null effects are simply a consequence of the size of the GIE (very small) and its high variability, as our meta-analysis indicates. On the other hand, we might also call the validity of our own results into question. This scenario seems more likely: the Bayesian analyses could not confirm that the GCE for the condition featuring only inanimate targets in Experiment 1 was robust; they could only confirm the robustness of the null effects for the rest of conditions. Also, if the GCE we found for bare nouns were reliable, it should have been observed by Finocchiaro et al. (2011), who obtained null results. Replication of our experiments is thus necessary. Importantly, it is worth noting that the "tendency" observed in our 50% condition of Experiment 1 and in Experiment 2, although not significant, was for incongruency (see Tables 17 and 20). A closer inspection of the data from the two experiments revealed considerable interindividual variability, sometimes depending on the gender of the target, and this may merit attention in future studies. This variability might explain why a GIE is sometimes observed, as in the experiments of Cubelli et al. (2005), Paolieri, Lotto, et al. (2010), and Paolieri et al. (2011). Hence, we must at least acknowledge the possibility that the answer to these differences between studies is neither the unreliability of the GIE nor that of the GCE, but the existence of individual differences within the population in the way that gender nodes compete. This requires further research (for a review on how individual differences may be key in explaining evidence in language processing, see Kidd et al., 2017). We will address this issue in more detail in the General Discussion. Finally, we might think that some kind of difference between European Portuguese and the other transparent Romance languages tested (Spanish and Italian) could be causing the direction of the effect to vary. Yet this seems highly unlikely, as seen in the previous analysis of European Portuguese and its similarities with Italian and Spanish (see subchapter 6.1). It is difficult to believe that, for such homogeneous group of languages, some show competition between lemmas based on gender, and another shows competition between gender nodes.

In sum, the present study provides supporting evidence for the idea that transparent Romance languages activate gender nodes during the lexical access of nouns in the absence of an agreement context, as predicted by the GSIH. Importantly, it can be seen as the first study to obtain evidence of the impact of animacy on the assessment of competitive gender effects

within the PWI paradigm. Animate target nouns showed null effects of gender and their presence within mixed stimuli lists blurred competitive gender effects between targets and distractors. We suggest two possible explanations for the results that should be tested in the future: the semantic prioritization of animates, and the attentional bias as defined by the Animate Monitoring Hypothesis. However, the validity of the whole study is at stake when assessing the reasons of observing the GCE, which would be theoretically and empirically expected but contradicts the evidence for Italian and Spanish, in which a GIE or null effects are obtained. Further research is hence needed here to clarify the robustness of the effect and the underlying mechanisms of grammatical gender processing. This is precisely one of the aims of our next set of experiments: to test the GCE obtained with bare nouns with a new sample of materials and a new group of Portuguese participants.

6.3 STUDY II: NOUN PHRASES AND BARE NOUNS PLUS GENDER TRANSPARENCY[37]

The present study aimed to further examine the representation and processing of grammatical gender during the production of nouns in European Portuguese, in order (1) to test the Late Selection Hypothesis, and (2) to attempt to replicate the GCE obtained in the first experiment of the previous study. Indeed, as seen in Chapter 5, the absence of the "determiner" congruency effect for Romance languages has been explained in terms of the late selection of the correct determiner form. More specifically, a mapping consisting of one-to-multiple gender-to-determiner forms in languages like Spanish or Italian entails a series of processes that must occur before the correct determiner within a specific gender value can be selected. Bürki et al.'s (2016) electroencephalographic PWI study with German participants and noun phrases suggested that the form of the noun (indicated by a phonological facilitation effect between target and distractor) is processed earlier than the determiner (indicated by the classical "GCE"). If their results are right, the computation of the forms of nouns is not one of the processes masking the "GCE" in Romance languages in comparison to Germanic and Slavic languages, because nouns' forms are always processed before determiners, even in German. The processes masking the "GCE" would therefore be: the computation of the stressing patterns in Spanish (to check whether the next noun is feminine and starts with a stressed "-a", in which case the

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

[37] Based on the paper: Ana Rita Sá Leite, Ângela Tomaz (*Universidad Nacional de Colombia*), Juan Andrés Hernández-Cabrera (*Universidad de La Laguna*), Isabel Fraga, & Montserrat Comesaña. *What a transparent Romance language with a Germanic gender-determiner mapping tells us about gender retrieval: Insights from European Portuguese* [Manuscript submitted for publication].

article to be selected is the usually masculine “el” and not the typical feminine “la”; “el agua” [F, the water]), and word retrieval and ordering in both Spanish and Italian (to check the existence of adjectives and their position within the noun phrase, because if they are to be produced and are prenominal, they may require a different article than that of the noun; “lo zaino” [the backpack] but “il grande zaino” [the big backpack]). These processes occur to make possible the selection of the proper determiner form within a specific gender value, and this masks any type of interaction that may have happened between determiner forms of different gender in an earlier phase of lexical access.

As seen in Chapter 5, French seems to be an exception among Romance languages. Similarly to Italian and Spanish, in French each gender value has more than one form of definite article: when the noun starts with a vowel, instead of the masculine “le” or the feminine “la”, the article that must be produced is “l’”. Foucart et al. (2010) propose that, since the choice between the normal and the shortened form does not depend on gender but merely on the noun’s form, “l’” does not have to be represented as a separate determiner, but as a simple reduced form obtained from a late phonetic adjustment. This would mean that competition between “le/la” would always occur, but full selection would take place after the form-based local context rule has been checked. Supposedly, this delay would make the process of determiner selection longer in French when compared to Germanic languages, but it would still be shorter than in certain Romance languages. Thus, perhaps the PWI paradigm can still be sensitive to competition between determiners in French. Foucart et al. (2010) tested a positive SOA that allowed interference between determiners to occur later in lexical access, and indeed found a GCE. However, Bürki et al. (2020) failed to replicate this, although they did obtain the effect twice with SOA 0 (simultaneous presentation of target and distractor) when considering both the naming times and the reading times of the distractor nouns according to previous measures of the participants’ reading performance. Hence, (1) the Late Selection Hypothesis seems to successfully accommodate the results on this matter when considering Italian and Spanish; (2) French may constitute a peculiar case, so that the role of the reading times investigated by Bürki et al. (2020) might be relevant not only to this issue but also in increasing the overall sensitivity of the PWI paradigm to detect certain effects of competition. We doubt, however, that the systematic absence of gender effects in Italian and Spanish when testing noun phrases can be explained by considering only reading times. This is so because such an absence of effects seems to be consequence of a more complex process of determiner selection, one that takes

USC UNIVERSIDADE DE SANTIAGO

places after a possible competition between gender nodes and between determiners of different gender. This complex process would mask these early competitive effects between target and distractor regardless of how much time the reading of the distractor takes.

Setting French aside, the possibility exists that the absence of effects in Spanish and Italian is a matter of another difference between language families (Germanic and Slavic vs. Romance minus French) or perhaps between languages of different gender transparency (opaque [including French] vs. transparent). Another possible explanation for the systematic null results observed in Italian and Spanish is that, if the GIE is indeed occurring, as our meta-analysis suggests, then this effect of competition between nouns of the same gender may somehow cancel the effect of competition between determiners of different gender.

As stated in Chapter 5, and also noted at the beginning of this chapter, European Portuguese is an interesting language for testing the Late Selection Hypothesis, in that it is a transparent Romance language, like Italian and Spanish, but has a one-to-one gender-determiner form mapping, like the Germanic and Slavic languages that we have tested thus far. Following the Late Selection Hypothesis, if the null results with noun phrases are a matter of mapping and not of any other differences between both groups of languages, a "GCE" (determiner congruency effect) should be obtained during the testing of native speakers of European Portuguese when they produce noun phrases (definite article + noun). Yet even if we obtain a "GCE" with noun phrases in European Portuguese, the current state of results does not allow us to claim that this is uniquely a reflection of competition between determiners, as the DSIH would hold, this because we obtained a GCE with bare nouns in Study 1.

Indeed, another of our aims in this study is to test the replicability of the GCE observed in the previous one. This effect contradicts previous evidence (the GIE), although it is in line with the cross-linguistic GCE, gender priming effects, and the organization of the syntactic properties that the classical models of language production propose (WEAVER++ and IN; i.e., nodes of different gender competing for selection without bidirectional connections between gender nodes and the nouns' lemma/lexeme). However, as already mentioned, the Bayesian analyses we conducted failed to confirm the robustness of our GCE, and we are still unsure about why we failed to obtain any effect in our Experiment 2 with only 25% of animate targets in the list (a similar percentage of that in Cubelli et al., 2005 and Paolieri, Lotto, et al., 2010; Paolieri et al., 2011). Besides, the null results obtained by Finocchiaro et al. (2011) remain an issue, and also contradict our GCE (although they are partially in line with a possible small,

USC UNIVERSIDADE DE SANTIAGO

heterogeneous and slippery GIE, one that can sometimes be hard to obtain, as our meta-analysis suggests).

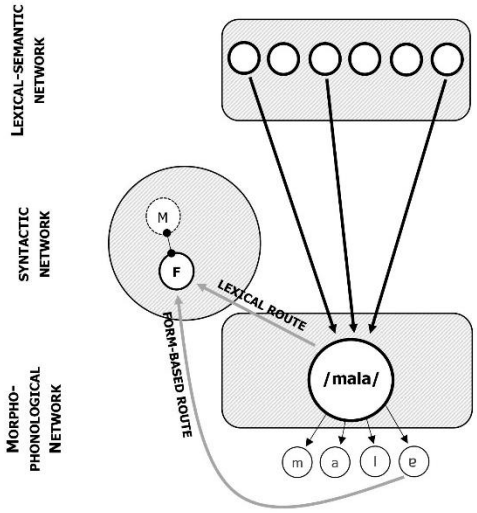
Apart from testing gender effects with noun phrases and bare nouns, we will also include in our design another variable that may be relevant to the debate: gender transparency. However, with the exception of the DSM model, which predicts a GIE and only explains the role of transparent cues on gender retrieval, leaving unanswered the question of how opaqueness is taken into consideration in this process, no other models of language production have assessed how gender would be retrieved in interaction with word form during lexical access. In fact, the proposal that has accumulated the greatest amount of evidence comes from language comprehension: Gollan and Frost's (2001) Dual-Route Model for gender retrieval. According to these authors, there are two available sources for the selection of gender values when reading or hearing a noun: (1) a lexical source based on abstract information stored within the noun, and (2) a non-lexical, form-based source that relies on morphological and phonological information. During the processing of transparent nouns our cognitive system can rapidly detect the phonological cues for gender via the form-based route. Then, checking processes take place to confirm that the selection of the gender value made through the form route matches the information obtained afterwards through the lexical route, as confirmed by Caffarra and colleagues (e.g., Caffarra et al., 2015; Caffarra & Barber, 2015). On the other hand, the selection of the gender value of opaque nouns has been shown to rely only on the lexical route, as the nominal endings are not identified as reliable cues for gender. Empirical ERP evidence consistently shows that effects of transparency are independent of the effects of agreement (e.g., Caffarra & Barber, 2015; Caffarra et al., 2015, but see Chapter 3 of this thesis) and that gender values are activated earlier for transparent than for opaque nouns (note that despite gender selection being more effective for transparent nouns, more cognitive resources are used in their processing, due to a larger amount of information being available; Caffarra & Barber, 2015; Caffarra et al., 2014). Importantly, transparent nouns have consistently shown a special robustness that protects them from inaccuracy and interference during gender retrieval, the opposite of what occurs with opaque nouns (Caffarra & Barber, 2015; Caffarra et al., 2014; Caffarra et al., 2015). The question of whether such routes are also the mechanisms of gender retrieval during lexical access in language production remains unanswered. We will hence propose an adaptation of the Dual-Route Model to language production, integrating those routes

USC UNIVERSIDADE DE SANTIAGO

in both the IN model or the WEAVER++ model (see Figures 14 and 15). We will then take this proposal as the basis for predicting the results of the present study.

Figure 14

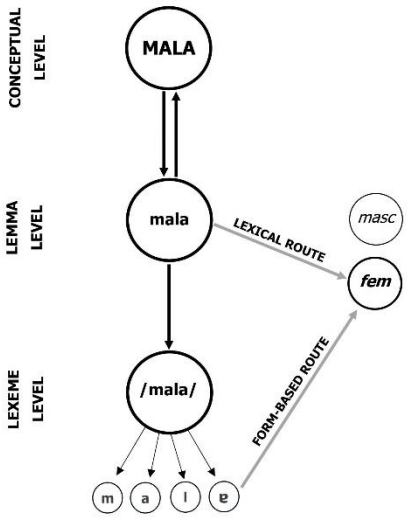
Representation of the routes of gender retrieval within the IN model



Note. The model is simplified. Also note that processing at the morpho-phonological network starts earlier than at the syntactic network, and hence the form-based route is expected to be activated earlier than the lexical route. In any case, we propose that the activation of both routes might overlap in time.

Figure 15

Representation of the routes of gender retrieval within WEAVER++



USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Note. The model is simplified. Also note that processing at the lemma level starts earlier than at the syntactic network, and hence the lexical route is activated earlier than the lexical route. In fact, we are unsure about whether or not the activation of both routes overlaps in time, and

within this model it is possible that the form-based route has a confirmatory role for the selection already carried out by the lexical route.

If the IN model is correct and Bürki et al.'s (2016) results with German speakers are generalizable, then the phonology of nouns would be processed in an earlier stage than grammatical gender. This would mean that the first source of gender retrieval would be the transparent cues present in the nominal endings of the nouns. Hence, once encoded, the presence of certain specific cues of transparency ("-o" and "-a") would directly activate a specific gender value – causing trouble in the processing of irregular nouns, but being extremely useful for the processing of transparent nouns. The gender value activated by the form-based cue would also receive activation from the lexical route (i.e., the gender value associated with that noun in our memory). Conversely, if we adopt the view of WEAVER++, and hence that the lemma is processed before the lexeme, then the retrieval of gender features will likely take place first through the lexical route. For transparency to have a role through a form-based route in the retrieval of gender in transparent nouns, we must either adopt an interactive view of our linguistic system, as proposed in Dell's account (1986, 1990), or use the self-monitoring internal speech system proposed by the authors of WEAVER++. In this way, the activation of the nominal ending of the lexeme and the activation of the gender node as stored in our memory would interact, contributing thus to the selection of that gender node. Alternatively, gender might already have been selected through the lexical route before the phonological form of the noun is completely encoded; but once it is encoded, gender cues could function as a confirmation of the gender retrieval that occurred at the previous stage – for irregular nouns, this would not be confirmed, and their processing would be more costly. This alternative idea might be a better fit for the feedback between lexeme and lemma accounted for by the self-monitoring system of WEAVER++. Finally, either if we conceive of noun form encoding starting before gender retrieval, or gender retrieval starting before noun form encoding, then transparent nouns will still rely on a more precise but more costly way of gender retrieval, capturing most of the available cognitive resources.

In sum, the present study aimed (1) to assess the retrieval of gender during noun phrase production in European Portuguese to test the Late Selection Hypothesis; (2) to test the replicability of the GCE obtained with bare nouns in our previous study; (3) to assess the role of ortho-phonological gender transparency in the effects. To do so, three PWI paradigm studies were carried out in which gender and nominal endings were manipulated in an orthogonal way

USC UNIVERSIDADE DE SANTIAGO

with noun phrases (Experiment 1) and bare nouns (Experiments 2 and 3). First, we predicted that a classical "GCE" would emerge in Experiment 1, this because according to the Late Selection Hypothesis European Portuguese should be the only transparent Romance language of those tested thus far to show a classic GCE with noun phrases containing definite articles. This GCE should reflect competition between determiners, or between determiners and gender values. In addition, following our previous results with bare nouns and solely inanimate stimuli, a GCE should also be observed in Experiments 2 and 3 with bare nouns as a naming instruction, even if an agreement context is not present. Taking gender transparency and our adaptation of the Dual-Route Model of gender retrieval to language production into account, we expected some type of impact of gender transparency on the GCE in all experiments. We proposed that transparent nouns rely either on an earlier source of information or on late confirmative extra feedback from the processing of the phonological cue at the lexeme level. On the contrary, opaque nouns rely only on the information stored in our memory – the lexical route. Therefore, if we were right, we expected: (1) opaque nouns to be more sensitive to disruptions during gender processing, and so, if the GCE is obtainable, it should also be especially substantial for opaque targets; and (2) the disruption caused by transparent nouns in the processing of other nouns to be greater than that caused by opaque nouns, due to the direct processing of gender through the form of the nouns and the higher amount of cognitive resources being used due to transparency and the matching between two routes. Note that interaction with transparency can occur with either a determiner or a genuine gender congruency effect (see Table 22 for the most likely outcomes in each experiment regarding gender congruency).

Table 22

Most likely outcomes of the experiments regarding the factor of gender congruency

	phrase type	**GCE**	**GIE**	**Support**
Experiment 1	noun phrase	Yes	NA	DSIH and GSIH
Experiment 2	bare noun	Yes	-	LSH
Experiment 3	bare noun	Yes	-	
Experiment 1	noun phrase	Yes	NA	DSIH
Experiment 2	bare noun	-	-	LSH
Experiment 3	bare noun	-	-	
Experiment 1	noun phrase	-	NA	None of the hypotheses
Experiment 2	bare noun	-	Yes	

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Experiment 3	bare noun	-	Yes	(against LSH – cancellation of effects "GCE" vs. GIE)
Experiment 1	noun phrase	-	NA	GSIH (against LSH)
Experiment 2	bare noun	Yes	-	
Experiment 3	bare noun	Yes	-	

Note. NA = Not applicable; LSH = Late Selection Hypothesis.

6.3.1 Experiment 1

6.3.1.1 Method

6.3.1.1.1 Participants

Seventy-six native speakers of European Portuguese studying at the University of Minho (69 female; M_{age} = 21.53 years, *SD* = 4.45) participated in the experiment. All were required to sign informed consent for experimentation with human subjects approved by the Ethics Committee of the University of Minho (SECSH 017/2015). Extra course credits were given to participants.

6.3.1.1.2 Materials

Sixty-four pictures were selected as targets from the International Picture Naming Project (IPNP) database (Szekely et al., 2004). Targets were distributed evenly in four categories according to their grammatical gender value and their ortho-phonological gender transparency: (a) 16 feminine and transparent words (e.g., "palmeira" [palmtree]), (b) 16 masculine and transparent words, (e.g., "barco" [boat]), (c) 16 feminine and opaque words (e.g., "ponte" [bridge]), and (d) 16 masculine and opaque words (e.g., "tapete" [rug]). Pictures from the four categories were matched through a one-way ANOVA with four conditions (masculine-transparent, feminine-transparent, masculine-opaque, feminine-opaque) for visual complexity (objectively defined by the digital size of the drawing), p = .299, and goodness-of-depiction (i.e., how well each picture illustrated the target nouns), p = .830, as obtained from the dataset in Szekely et al. (2005). Target words were matched through the same one-way ANOVA across categories on a per-million frequency (Log10), this obtained from the SubtLex-PT database (Soares et al, 2015). They were also matched on the number of phonological and orthographic neighbours (N), word length (in letters), and mean logarithmic bigram frequency, this obtained from the P-PAL database (Soares, Iriarte, et al., 2018). Finally, they were controlled for

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

imageability, concreteness, and subjective frequency, this obtained from the Minho Word Pool database (MWP; Soares et al., 2017). Note the mean values of every controlled variable per target type in Table 23 (all *ps* > .110).

Table 23

Means and standard errors of the controlled variables per target type for the 64 experimental pictures in Experiment 1

	Feminine Transparent	Masculine Transparent	Feminine Opaque	Masculine Opaque
FpM	3.10 (.15)	3.40 (.12)	3.08 (.16)	2.89 (.13)
PN	2.50 (.47)	5.19 (1.72)	3.19 (1.03)	2.31 (.95)
ON	2.75 (.59)	5.63 (1.49)	3.25 (1.00)	2.44 (1.08)
L	6.69 (.31)	6.31 (.53)	5.75 (.40)	6.06 (.38)
MLBF	2.55 (.11)	2.81 (.29)	2.45 (.14)	2.42 (.14)
VC	17,265.13 (1,891.93)	14,295.12 (1,407.58)	18,067.06 (2,590.53)	13,230.19 (2,210.94)
GoD	5.88 (.11)	5.96 (.09)	5.81 (.23)	5.76 (.68)
I	5.91 (.08)	6.15 (.07)	5.97 (.08)	5.96 (.09)
C	6.52 (.06)	6.55 (.05)	6.27 (.15)	6.44 (.08)
SF	5.10 (.31)	5.13 (.32)	4.63 (.34)	4.39 (.32)

Note. FpM = Frequency per million (Log_{10}). Obtained from the SubtLex-PT database (Soares et al., 2015). PN = Phonological number of neighbours; ON = Orthographic number of neighbours; L = Length (in letters); MLBF = Mean log bigram frequency. All obtained from the P-PAL database (Soares, Iriarte, et al., 2018). VC = Visual complexity; GoD = Goodness-of-depiction. Both obtained from the International Picture Naming Project (IPNP) database (Szekely et al. 2004). I = Imageability; C = Concreteness; SF = Subjective Frequency. All obtained from the Minho Word Pool database (Soares et al., 2017).

For each picture, four European Portuguese distractor nouns were selected (256 in total, 64 per condition); two of these had the same grammatical gender, and the other two had different grammatical gender. All target and distractor nouns were inanimate and thus had grammatical gender, a measure of control justified in our first experiment (see the previous subchapter). Note that none of these distractors showed any semantic relationship with the targets with which they were paired, as they were from different semantic categories, nor was there any significant

USC DE SANTIAGO DE COMPOSTELA

orthographic overlap between them, according to the NIM database (Guasch et al., 2013). In addition, all targets and their respective distractors differed in their initial phoneme to avoid any kind of facilitation during the naming task (Bi et al., 2009).

Targets were assigned to four experimental conditions as a function of their relation in terms of Gender Congruency (GC) and Transparency Congruency (TC) with the distractors. For example, with the feminine transparent target "palmeira" (palmtree), four distractors were selected: a) congruent in gender and in transparency category (GCTC), such as "carroça" (carriage); b) incongruent in gender but congruent in transparency (GITC), such as "ginásio" (gymnasium); c) congruent in gender but incongruent in transparency (GCTI), such as "grafite" (graphite); and d) incongruent in both gender and transparency category (GITI), such as "foguete" (rocket). The same conditions were created for feminine opaque targets as for transparent and opaque masculine ones.

Distractors were matched through a one-way ANOVA across 16 conditions per lexical frequency (log10), this obtained from the SubtLex-PT database (Soares et al., 2015), on the number of phonological and orthographic neighbours (N), word length (in letters), and mean logarithmic bigram frequency, this obtained from the P-PAL database (Soares, Iriarte, et al., 2018; see mean values per target type in Table 24a and Table 24b, all *ps* > .062).

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Table 24a

Means and standard errors of the controlled variables for distractors per experimental condition of feminine targets in Experiment 1

	FT GCTC	FT GCTI	FT GITC	FT GITI	FO GCTC	FO GCTI	FO GITC	FO GITI
FpM	2.63 (.19)	2.33 (.25)	2.56 (.17)	2.55 (.19)	2.68 (.30)	2.98 (.25)	2.75 (.21)	2.56 (.20)
PN	2.5 (.54)	2.56 (.57)	2.19 (.63)	2.50 (.58)	2.13 (.86)	3.44 (.93)	2.38 (.89)	2.56 (.87)
ON	2.94 (.57)	2.25 (.57)	2.44 (.71)	2.19 (.47)	2.13 (.68)	3.06 (.84)	2.63 (.92)	2.25 (.61)
L	6.38 (.30)	6.44 (.35)	6.50 (.27)	6.31 (.31)	6.13 (.43)	6.00 (.29)	6.06 (.44)	6.38 (.30)
MLBF	2.54 (.07)	2.55 (.09)	2.58 (.10)	2.54 (.10)	2.52 (.13)	2.60 (.09)	2.47 (.08)	2.49 (.12)
OVTD	.16 (.01)	.14 (.03)	.10 (.02)	.14 (.03)	.14 (.03)	.12 (.02)	.16 (.03)	.12 (.03)

Note. FT and FO = Feminine Transparent and Feminine Opaque. GCTC, GCTI, GITC and GITI = Gender Congruent and Transparency Congruent, Gender Congruent and Transparency Incongruent, Gender Incongruent and Transparency Congruent, and Gender Incongruent and Transparency Incongruent, respectively. FpM = Frequency per million (Log_{10}). Obtained from the SubtLex-PT database (Soares et al., 2015). PN = Phonological number of neighbours; ON = Orthographic number of neighbours; L = Length (in letters); MLBF = Mean log bigram frequency. All obtained from P-PAL database. OVPN = Orthographic overlap between targets and distractors obtained from the NIM database (Guasch et al., 2013).

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Table 24b

Means and standard errors of the controlled variables for distractors per experimental condition of masculine targets in Experiments 1 and 2

	MT GCTC	MT GCTI	MT GITC	MT GITI	MO GCTC	MO GCTI	MO GITC	MO GITI
FpM	2.79 (.21)	2.87 (.19)	2.80 (.14)	2.66 (.13)	2.30 (.20)	2.62 (.19)	2.21 (.17)	2.60 (.15)
PN	3.75 (.99)	3.63 (1.08)	4.50 (1.14)	3.31 (1.16)	1.69 (.59)	1.81 (.44)	2.13 (.73)	2.13 (.60)
ON	4.38 (1.29)	3.00 (1.02)	4.31 (1.05)	2.81 (.98)	1.38 (.55)	1.50 (.67)	1.63 (.49)	2.56 (.86)
L	6.13 (.36)	6.00 (.35)	5.94 (.39)	6.63 (.44)	6.25 (.36)	6.44 (.24)	6.81 (.34)	6.13 (.30)
MLBF	2.57 (.09)	2.55 (.09)	2.60 (.12)	2.56 (.10)	2.13 (.06)	2.28 (.12)	2.52 (.12)	2.35 (.09)
OVTD	.16 (.01)	.08 (.02)	.06 (.03)	.10 (0.2)	.15 (.03)	.16 (.03)	.19 (.03)	.14 (.04)

Note. MT and MO = Masculine Transparent and Masculine Opaque. GCTC, GCTI, GITC and GITI = Gender Congruent and Transparency Congruent, Gender Congruent and Transparency Incongruent, Gender Incongruent and Transparency Congruent, and Gender Incongruent and Transparency Incongruent, respectively. FpM = Frequency per million (Log_{10}). Obtained from the SubtLex-PT database (Soares et al., 2015). PN = Phonological number of neighbours; ON = Orthographic number of neighbours; L = Length (in letters); MLBF = Mean log bigram frequency. All obtained from P-PAL database. OVPN = Orthographic overlap between targets and distractors obtained from the NIM database (Guasch et al., 2013).

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Finally, eight different pictures from the IPNP database were selected as training items, along with eight new distractors. Four different lists were created for counterbalancing purposes. Each of these featured the same four conditions based on gender and transparency congruency, and included the 64 target pictures (16 per condition). Half of the experimental items were masculine (32) and half were feminine (32). Likewise, half of the masculine and feminine pictures had a transparent ending (16) and half had an opaque ending (16). The same structure on gender value and transparency category was also followed for the distractors. Thus, each target picture was presented four times, one time per list and each time associated with a different distractor. Participants were randomly assigned to one of the four lists, assuring the same number of participants per list.

6.3.1.1.3 Procedure

Participants were tested individually in a soundproof booth. As in previous studies (e.g., Cubelli et al., 2005), each participant was first familiarized with the set of pictures to guarantee high naming agreement scores in the experimental task. To this end, all pictures were presented in single slides, along with the nouns that participants had to use to name them in the task. Each picture and corresponding noun were presented for 2 s and no overt response was required. The pictures were presented using Microsoft PowerPoint 2010. Words were shown in lowercase letters (Agency FB 36-point) below the respective picture. Stimuli presentation was randomized across participants.

In the PWIP, participants were instructed to name each picture using a noun phrase (an article followed by a noun, e.g., “a palmeira” [the palmtree]) as quickly and accurately as possible, while ignoring the distractor noun superimposed onto the picture. All the pictures were presented at the centre of the computer screen. The distractors, presented at the centre of the pictures with a 0 ms SOA, were shown in lowercase letters in Courier New font, 14-point (see Cubelli et al., 2005 for a similar procedure). Each experimental trial had the following structure: a fixation point (+) at the centre of the computer screen, for 500 ms; the target picture with the superimposed distractor for 2,000 ms or until response; a blank space for 500 ms as an inter-trial interval. Trials were presented randomly per participant with a pause after the first 32 trials. Response latencies were measured from the onset of the stimulus to the beginning of the naming response. Stimuli presentation was done using the DMDX software (Forster & Forster, 2003). Naming RTs were recorded by the voicekey from the presentation of the target

USC UNIVERSIDADE DE SANTIAGO

to the onset of the naming response and then checked offline using the CheckVocal software (Protopapas, 2007). The session lasted approximately 20 min.

6.3.1.2 Results and interim discussion

Data corresponding to incorrect responses (naming errors [3.31%] and non-responses [2.12%]) were not included in the analysis. The pictures "câmara" (camera) and "abóbora" (pumpkin) were excluded, since they were incorrectly recognized by at least one third of the participants (33.33%). Importantly, the masculine distractor "vinho" (wine), belonging to the second list and paired with feminine target "ponte" (bridge), was incorrectly classified as gender congruent, and thus in the analysis the condition of gender congruency lost one item per participant and the condition of gender incongruency gained one on the second list.

In addition, RTs that were more than 2.5 *SD*s above or below the mean for each participant in each condition were replaced by the mean response latency for the respective participant (3.03% in total). Mean response latencies (in milliseconds) per condition are presented in Table 25. As in previous studies (Bordag & Pechmann, 2008; Costa et al., 1999; Cubelli et al., 2005; Paolieri, Lotto, et al., 2010; Paolieri et al., 2011), accuracy was not analysed due to the absence of gender-agreement errors (none of the participants used a masculine article for a feminine target or vice versa).

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Table 25

Mean response latencies (in milliseconds, standard deviations in parentheses) for Experiment 1 (noun phrases)

	Target type							
	Feminine transparent		Masculine transparent		Feminine opaque		Masculine opaque	
Gender	Congruent	Incongruent	Congruent	Incongruent	Congruent	Incongruent	Congruent	Incongruent
Transparency congruency	867 (189.55)	861 (188.04)	840 (187.42)	836 (179.97)	843 (203.45)	847 (199.25)	848 (194.89)	838 (173.36)
GCE	-6		-4		4		-10	
Transparency incongruency	860 (175.42)	874 (188.14)	829 (165.07)	818 (167.66)	822 (170.44)	873 (201.28)	835 (189.19)	857 (200.12)
GCE	14		-11		51		22	

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

We analysed RTs for target naming with linear mixed-effects (lme) models. The lme on RTs were conducted with participants and items as crossed random factors. The models were fit using the lme4 R package (Bates et al., 2015) as well as the LMerTest (Kuznetsova et al., 2017) and the lsmeans package (Lenth, 2016). This allowed us to contrast simple effects with differences of the least squares means. There was no averaging of the data prior to the analyses. We investigated a 2x2x2x2 interaction between the design factors Target Gender (feminine, masculine), Target Transparency (transparent, opaque), Gender Congruency (GC: congruent [GC], incongruent [GI]) and Transparency Congruency (TC: congruent [TC], incongruent [TI]). Because of the strong non-normality of residuals, a box-cox transformation of RT was performed with a power transformation of -0.22 (Box & Cox, 1964). Following Barr and colleagues (2013), a maximum random-effects model was carried out with random slopes for all factors by participant, plus Gender Congruency and Transparency Congruency (factors within by items) by item. This model shows perfect collinearity (correlation of 1 and -1 between random intercepts and slope for both random parts of the model) resulting in serious convergence problems and negative eigenvalues. The final estimated model that reached convergence had random intercepts for participants and items and random slopes only for Transparency Congruency by participants and Gender Congruency by items: model.p='RT.power ~ Target_Gender*GC*NEC*Transparency + (NEC|Participant) + (GC|Target)' (NEC means Transparency Congruency)

Importantly, in order to assess the predictive power of item properties on the RTs, a stepwise linear regression was carried out with nine items' covariate variables as predictors (the number of phonological and orthographical neighbours, length [number of letters], and the mean log bigram frequency of the distractors, as obtained through the P-PAL database; the orthographic overlap between targets and distractors as obtained through the NIM database; the degree of imageability, concreteness, and subjective frequency of the distractors as obtained from the Minho Word Pool database; and the logarithmic frequency per million as obtained by the SubtLex-PT database). None of the nine candidate variables was significant.

The results of the RTs revealed a significant two-fold interaction between Gender Congruency and Target Transparency, $F(1, 54.9) = 5.51$; $p = .02$. Post-hoc comparisons revealed that the effect of GC (GC < GI) was mainly due to opaque targets (GC opaque targets = 837 ms; GI opaque targets = 854 ms, $p = .003$). The main effect of gender congruency was marginally significant $F(1, 54.9) = 3.44$; $p = .069$, as gender congruent target-distractor pairs

USC UNIVERSIDADE DE SANTIAGO

were named faster than gender incongruent ones (GC < GI, 843 and 850 ms, respectively). Another two-fold interaction approached significance, Gender Congruency x Transparency Congruency $F(1, 4241.5) = 3.32$; $p = .068$. Planned comparisons revealed that the GCE was due mainly to transparency incongruent distractors (when target and distractor were from different categories of transparency, GC = 836 ms, GI = 855 ms; for instance, an opaque target such as "*tapete*" (rug) paired with a transparent distractor such as "*xadrez*", $p = .0095$; conversely, for targets and distractors of the same transparency category, $p = .90$).

In short, as predicted by the Late Selection Hypothesis (Miozzo & Caramazza, 1999), a GCE was obtained for the first time in a PWI paradigm study with speakers of a Romance language producing noun phrases (determiner + noun), hence confirming that European Portuguese indeed behaves like an early-selection language and hence the body of results is a matter of gender-determiner form mapping. However, the GCE in European Portuguese is not as simple as the one reported for Germanic and Slavic languages. Our orthogonal design allowed us to obtain evidence supporting the notion that transparency matters in language production: the GCE interacted with the transparency category of the targets. More specifically, the GCE was due mainly to opaque targets. Furthermore, planned comparisons revealed that the effect tended to be more robust for targets paired with distractors of another transparency category. As displayed in the mean RTs of Table 24, the gender effect sizes for the transparency incongruent conditions, especially for opaque targets, are indeed the largest. This is not surprising. These results are in line with our adaptation of Gollan and Frost's Dual-Route Model (2001) to language production, and also with the data reported in the electrophysiological literature on gender-transparency processing during language comprehension (Caffarra et al., 2015). It seems that the use of different routes to retrieve gender causes interference between targets and distractors. Besides, when distractors are transparent nouns, (1) the use of the form-based route allows for an earlier activation of gender, and (2) the greater amount of available information to encode gender demands more cognitive resources to effectively select the gender value. This may cause a greater disruption in the gender selection of opaque targets when distractors are transparent than vice-versa. We will explain these issues in more detail in the General Discussion. Importantly, taken alone, the results of this experiment are insufficient to allow for us to decide whether this GCE is a sole reflection of determiner competition or also of gender competition (DSIH vs. GSIH). The following two experiments will address this issue by asking participants to use only bare nouns to name the pictures.

USC UNIVERSIDADE DE SANTIAGO

6.3.2 Experiment 2

6.3.2.1 Method

6.3.2.1.1 Participants

Eighty native speakers of European Portuguese studying at the University of Minho (72 female; M_{age} = 21.53 years, *SD* = 4.68) from the same population as those in Experiment 1 participated in this experiment. Participants were required to sign informed consent for experimentation with human subjects approved by the Ethics Committee of the University of Minho (SECSH 017/2015). Extra course credits were given to participants.

6.3.2.1.2 Materials

We used the same materials as in Experiment 1.

6.3.2.1.3 Procedure

We followed the same procedure used in Experiment 1, with the exception that participants were required to produce bare nouns (without articles).

6.3.2.2 Results and interim discussion

Data corresponding to incorrect responses (naming errors [2.50%] and non-responses [3.43%]) were not included in the analysis. The pictures "câmara" (camera) and "abóbora" (pumpkin) were again excluded for being incorrectly recognized by at least one third of the participants (33.33%). Given that the same materials as those from Experiment 1 were used, the condition of GC had one less item per participant and the condition of GI had one extra item on the second list due to an incorrect categorization (distractor "vinho" with target "ponte"). Again, RTs that were more than 2.5 SDs above or below the mean for each participant in each condition were replaced by the mean response latency for the respective participant (2.88% in total). Mean response latencies (in milliseconds) per condition are presented in Table 26. As in Experiment 1, accuracy was not analysed due to the absence of gender-agreement errors.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Table 26

Mean response latencies (in milliseconds, standard deviations in parentheses) for Experiment 2 (bare nouns)

	Target type							
	Feminine transparent		Masculine transparent		Feminine opaque		Masculine opaque	
Gender	Congruent	Incongruent	Congruent	Incongruent	Congruent	Incongruent	Congruent	Incongruent
Transparency congruency	894 (187.36)	898 (189.59)	878 (181.06)	859 (171.52)	852 (183.89)	882 (211.22)	874 (196.91)	876 (207.77)
GCE	4		-19		30		2	
Transparency incongruency	884 (188.01)	901 (194.78)	873 (182.98)	851 (169.84)	865 (199.01)	878 (195.88)	887 (192.31)	895 (195.64)
GCE	17		-22		13		8	

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

The same analysis with the same factors included in Experiment 1 was replicated here. Because of the strong non-normality of residuals, a box-cox transformation of RT was performed with a power transformation of -1 (Box & Cox, 1964). Following Barr and colleagues (2013), a maximum random-effects model was carried out with random slopes for all factors by participants and Gender Congruency and Transparency Congruency (factors within by items) by items. Because of non-convergence problems and negatives eigenvalues, the confidence intervals of random effects were checked graphically (Pinheiro & Bates, 2000). The final estimated model had random intercepts for participants and items, and random slopes for Target Gender and Transparency Congruency by participants and Gender Congruency by items:

model.p='RT.p ~ Target_Gender*GC*NEC*Transparency + (Target_Gender + NEC|Participant) + (GC|Target)' (NEC means Transparency Congruency)

As in Experiment 1, a stepwise regression with RTs as outcome and the same nine items' covariate variables as predictors was carried out. None of the candidate covariates variables was significant.

The results for the RTs revealed non-significant effects involving the factor Gender Congruency, except for a marginally significant two-fold interaction between Target Gender and Gender Congruency $F(1, 57.3) = 3.61$; $p = .062$. However, planned comparisons failed to show significant interactions. On the other hand, a significant two-fold interaction between Transparency Congruency and Target Transparency was observed, $F(1, 4404) = 6.96$; $p = .008$. Planned comparisons revealed that when targets were opaque, they were named faster when paired with distractors of the same transparency category (also opaque) than when paired with transparent distractors (for opaque targets, TC < TI, 871 ms and 881 ms, respectively, $p = .03$).

The absence of gender effects (GIE or GCE) with bare nouns fits the null results obtained by Finocchiaro et al. (2011) for Spanish and Italian. Importantly, they support the interpretation of the GCE of our first experiment as a reflection of competition between determiners and ultimately the DSIH. However, although there were no significant effects of GC, there was an effect of transparency that might suggest that although pure gender effects do not emerge in the RTs, gender nodes are somehow activated during bare noun production, creating different degrees of activation in different gender nodes and inducing disruption on lexical access. More specifically, opaque nouns here benefited from being paired with distractors of the same category of transparency (also opaque). The size of the effects reported in the present study

USC UNIVERSIDADE DE SANTIAGO

show a trend in which the GCE would be greater for opaque targets, in line with the results obtained in Experiment 1. This leads us to believe that Experiment 2 reflects a similar pattern of results as those reported in Experiment 1, supporting the GSIH. Because of this, and also because opaque targets seemed to somehow have a special role in both experiments, we decided to conduct an additional experiment with bare nouns (Experiment 3) using a simpler design. In this experiment, we used only opaque targets (i.e., the factor Target Transparency was dropped) with a partial new set of materials. This time, we had four lists, but each participant completed every list at different points in time, which allowed us to obtain many more entries per stimuli. In addition to the simpler design, this increased our statistical power. Importantly, to increase the similarity of our experiments to those that obtained a GIE with Romance languages (Cubelli et al, 2005; Paolieri, Lotto, et al., 2010; Paolieri et al., 2011), we included a similar percentage of animate stimuli (28% of target nouns; 9% of distractor nouns) within the list, as this was the main differential factor in the control of the stimuli between our experiments and theirs.

6.3.3 Experiment 3

6.3.3.1 Method

6.3.3.1.1 Participants

Forty-eight native speakers of European Portuguese studying at the University of Minho (45 female; M_{age} = 20.48 years, *SD* = 2.70) from the same population as previous experiments participated in this experiment. Participants were required to sign informed consent for experimentation with human subjects approved by the Ethics Committee of the University of Minho (SECSH 017/2015). Extra course credits were given to participants.

6.3.3.1.2 Materials

We selected 32 opaque pictures as targets, including 22 of those used in Experiments 1 and 2. The procedure was the same as that in the previous experiments. However, we also selected 32 items as transparent fillers. Every picture was presented four times, each time in a different list with a different distractor (see Tables 5 and 6 for the control on targets and distractors). With the aim of increasing the stimuli pool, this time 13% of the pictures and distractors were animate. Furthermore, in this way we were able to emulate previous studies in which 8 to 33% of animate stimuli were included, and effects were systematically obtained with bare nouns (e.g., Cubelli et al., 2005; Paolieri, Lotto, et al., 2010; Paolieri et al., 2011).

USC UNIVERSIDADE DE SANTIAGO

However, instead of having a different set of participants per list, each participant saw all the lists, as was done in previous studies (Cubelli et al., 2005; Finocchiaro et al., 2011; La Heij et al., 1998; O'Rourke, 2007; Paolieri, Lotto, et al., 2010; Paolieri et al., 2011). This allowed us to increase statistical power. The order of presentation was counterbalanced across participants.

Masculine and feminine target pictures and nouns were matched through a one-way ANOVA for visual complexity and goodness-of-depiction, as obtained from the dataset in Szekely et al. (2005), per-million frequency (Log10), this taken from the SubtLex-PT database (Soares et al, 2015), number of phonological and orthographic neighbours (N), word length (in letters), and mean logarithmic bigram frequency, this obtained from the P-PAL database (Soares, Iriarte, et al., 2018), and finally, imageability, concreteness, and subjective frequency for the 78% of available target nouns in the Minho Word Pool database (MWP; Soares et al., 2017). All *ps* > .121 (see mean values in Table 27). Distractor nouns were matched for the same variables except the picture-related ones (visual complexity and goodness-of-depiction). All *ps* > .178 (see mean values in Table 28). One distractor was missing from the SubtLex-PT database (127/128 distractor words were included in the analyses). Analyses for imageability, concreteness, and subjective frequency were carried out for 58% of the distractor nouns, those available in the Minho Word Pool database. Although concreteness showed significant differences in the ANOVA (p = .020), post-hoc analyses showed no significant pair-wise comparisons (p > .191).

Table 27

Means and standard errors (SE) of the controlled variables per target type for the 32 experimental stimuli in Experiment 3

	Feminine	Masculine
FpM	3.13 (.18)	3.37 (.19)
PN	2.81 (1)	3.38 (.935)
ON	5.81 (.40)	5.56 (.44)
L	2.88 (1)	3.56 (1.09)
MLBF	2.48 (.14)	2.55 (.13)
VC	18,502.62 (2,507.28)	16,728.79 (2,523.20)
GoD	5.78 (.23)	5.91 (.16)
I	5.97 (.08)	6.10 (.07)

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

C	6.24 (.14)	6.49 (.07)
SF	4.63 (.34)	4.77 (.27)

Note. FpM = Frequency per million (Log_{10}). Obtained from the SubtLex-PT database (Soares et al., 2015). PN = Phonological number of neighbours; ON = Orthographic number of neighbours; L = Length (in letters); MLBF = Mean log bigram frequency. All obtained from the P-PAL database (Soares, Iriarte, et al., 2018). VC = Visual complexity; GoD = Goodness-of-depiction. Both obtained from International Picture Naming Project (IPNP) database (Szekely et al. 2004). I = Imageability; C = Concreteness; SF = Subjective Frequency. All obtained from the Minho Word Pool database, Soares et al., 2017).

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Table 28

Means and standard errors (SE) of the controlled variables for distractors per experimental condition of masculine targets in Experiment 3

	F GCTC	F GCTI	F GITC	F GITI	M GCTC	M GCTI	M GITC	M GITI
FpM	2.66 (.29)	2.96 (24)	2.58 (.26)	2.75 (.17)	2.57 (.23)	2.68 (.25)	2.41 (.23)	2.74 (.22)
PN	1.81 (.84)	3.50 (1.13)	2.69 (.93)	2.44 (.90)	2.81 (.90)	3.81 (1.06)	2.56 (.58)	2.75 (.95)
ON	1.94 (.65)	3 (1.07)	2.44 (.68)	2.06 (.61)	3.06 (.92)	3.81 (.89)	2.00 (.51)	3.38 (1.09)
L	6.13 (.41)	6.00 (.30)	5.94 (.44)	6.25 (.30)	5.81 (.44)	6.06 (.36)	6.06 (.41)	6.13 (.36)
MLBF	2.49 (.13)	2.58 (.10)	2.45 (.13)	2.52 (.09)	2.42 (.14)	2.62 (.15)	2.45 (.10)	2.54 (.06)
OVTD	.16 (.03)	.10 (.03)	.10 (.03)	.08 (.03)	.034 (.03)	.12 (.09)	.16 (.03)	.17 (.03)

Note. FT and FO = Feminine Transparent and Feminine Opaque. GCTC, GCTI, GITC and GITI = Gender Congruent and Transparency Congruent, Gender Congruent and Transparency Incongruent, Gender Incongruent and Transparency Congruent, and Gender Incongruent and Transparency Incongruent, respectively. FpM = Frequency per million (Log_{10}). Obtained from the SubtLex-PT database (Soares et al., 2015). PN = Phonological number of neighbours (N); ON = Orthographic number of neighbours (N); L = Length (in letters); MLBF = Mean log bigram frequency. All obtained from P-PAL database. OVPN = Orthographic overlap between targets and distractors obtained from the NIM database (Guasch et al., 2013).

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

6.3.3.1.3 Procedure

We followed the procedure used in Experiment 2, hence participants were required to produce bare nouns (without the use of articles). In addition, each list was presented on a different day (with a time interval of between two and three days) to avoid any biases related to residual memory effects.

6.3.3.2 Results

Data corresponding to incorrect responses (naming errors and non-responses) were not included in the analysis (3.34%). Also, RTs that were more than 2.5 *SD*s above or below the mean for each participant in each condition were removed (2.76% of the total). Mean response latencies are presented in Table 29.

Table 29

Mean response latencies (in milliseconds, standard deviations in parentheses) for Experiment 3 (bare nouns and opaque targets)

	Target type			
	Feminine opaque		Masculine opaque	
Gender Transparency	Congruent	Incongruent	Congruent	Incongruent
congruency	807 (169.29)	822 (177.59)	797 (165.84)	792 (143.95)
GCE	15		-5	
Transparency incongruency	819 (167.20)	826 (172.41)	802 (159.91)	789 (157.04)
GCE	7		-13	

We analysed RTs (power transformed with a lambda of -0.5, Box & Cox, 1964) for target naming with linear mixed effects (lme) models. As in the two previous experiments, a maximum random-effects model was carried out with random slopes for all factors by participants and Gender Congruency (factor within by items) by items. Because of collinearity and convergence problems, the final model had a random slope only for the GC factor by item. model.p='RT.p ~ Target_Gender*GC*NEC + (NEC|Participante) + (GC|Target)' (NEC = Transparency Congruency)

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

The models were fit using the lme4 R package (Bates et al., 2015) as well as the LMerTest (Kuznetsova et al., 2017) and the lsmeans package (Lenth, 2016). This allowed us to contrast simple effects with differences of least squares means. There was no averaging of the data prior to the analyses. The design was the same as in previous experiments, except for the absence of the factor Target Transparency (i.e., it was a 2x2x2 design). The stepwise regression with RT as the outcome and the nine items characteristics as predictors did not yield any significant predictive model. Although the pattern of data was similar to that of Experiment 2 (especially for feminine targets), no significant effects were obtained.

6.3.4 Discussion

In the present study, we sought to explore grammatical gender processing in European Portuguese through three different PWI tasks. The first required participants to name the pictures using noun phrases formed by definite articles and nouns, and the other two required bare nouns. Our design considered not only the gender congruency between target and distractor, but also the specific gender of the target and the gender transparency of targets and distractors. Following the tenets of the Late Selection Hypothesis (Miozzo & Caramazza, 1999), because European Portuguese has a one-to-one mapping between gender and determiners, we expected to obtain a classical "GCE" in the first experiment, that is, a determiner congruency effect according to the DSIH. Since a genuine GCE had previously been obtained in Study 1 when testing bare nouns in European Portuguese, this "GCE" might also have been reflecting genuine gender competition according to the GSIH. In order for such support to be found for the GSIH, a GCE should have been obtained in the second and third experiments. Also, in terms of transparency, we predicted that opaque target nouns would be more prone to suffer from interference during gender selection, in that they rely on lesser sources of information to retrieve a gender value. Likewise, we expected transparent distractor nouns to cause more interference than opaque ones. The GCE, understood either as a determiner (Experiment 1) or a genuine gender effect (all experiments), should hence have been greater for opaque targets, and for targets paired with transparent distractors.

Yet what we obtained was our second most likely outcome (see Table 22): the results showed a classical "GCE" in Experiment 1 in interaction with transparency, but no effects of gender competition in Experiment 2 and 3. This ultimately supports the LSH and seems to be in agreement with the DSIH, which understands the "GCE" as a result of competition between

USC UNIVERSIDADE DE SANTIAGO

determiner forms. Importantly, the "GCE" was due to opaque target nouns, revealing that this effect is not as simple as the one obtained in opaque languages. In addition, a marginal interaction with the factor Transparency Congruency in this experiment revealed that distractors of a different category of transparency than that of the target seemed to be interfering more than those of the same category. An inspection of the mean RTs reveals that this is especially true for opaque targets paired with transparent distractors (mean GCE of 37.5 ms). This is in line with our adaptation of the Dual-Route Model for language production, according to which more interference is expected to occur during the selection of a gender value for opaque targets, especially when these are paired with transparent distractor nouns. More specifically, transparent distractor nouns unequivocally activate a certain gender node through the form-based route, in addition to the activation coming also from their own lexical route. However, opaque distractor nouns activate gender exclusively through the lexical route, and hence the amount of activation is lower, and generates less interference in gender selection than a transparent distractor noun. In fact, transparent nouns have been shown to consume more cognitive resources (see Caffarra & Barber, 2015; Caffarra et al., 2014, 2015). Ultimately, this transparency interference in the selection of the proper gender value is reflected in the selection of the correct determiner. This suggests that activation cascades from gender values to determiners, which is actually hypothesized in the *Primed Unitized Activation Hypothesis* and is shown in the singular-plural paradigm when testing the selection of determiner forms for plural and diminutive nouns (see Alario & Caramazza, 2002). Note that if the gender selection of nouns was resolved before the determiner selection, we would not expect that the activation would come in different amounts and times from the different routes of gender retrieval to influence the competition between determiners of different gender. Interaction between both effects thus happens because both processes overlap in time due to a cascaded flow of activation. The differences in the amount of activation reaching one gender node or another are then reflected in the amount of activation distributed across determiner forms, hence creating different patterns of activation and competition depending on which and how many routes are sending activation. Once gender selection finishes, different amounts of activation have already reached each determiner representation, influencing how the process of determiner selection takes place.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

In Experiment 2, even though we did not find effects of gender congruency, opaque target nouns also benefited from being paired with other opaque distractors in comparison to

transparent distractors. This means that the activation coming from different routes of gender are creating some kind of disruption in the processing of the target's gender. This could be understood as evidence for either competition between gender values (supporting the GSIH), or as evidence for the idea that the amount of available resources when a transparent distractor noun is being processed are insufficient for the processing of an opaque noun, which is a more likely interpretation given the data we have (i.e., the routes of transparency are activating gender differently, but gender does not seem to be competing). It is also worth noting that, although this effect of transparency might be seen as a mere effect of form (i.e., of phonological facilitation of nominal endings), we do not think that this is the case. In this experiment, 36 of 64 of the distractors had the same nominal ending as their associated opaque target noun (mostly "-e"). Nevertheless, we need to bear in mind here that there is one condition with a perfect match of the coinciding final letters in targets and distractors: transparent targets paired with transparent distractors that are gender congruent (feminine targets ending in "-a" with distractors ending in "-a"; masculine targets ending in "-o" with distractors ending in "-o"). There were 32 gender congruent transparent targets paired with transparent distractors, and hence, 32/32 sharing the final letter with the distractor; conversely, there were 32 gender incongruent transparent targets paired with transparent distractors, and hence, 0/32 sharing the final letter with the distractor. If the coincidence between the final letter of the target and that of the distractor were of such relevance, we would have expected to find some kind of phonological facilitation effect for the former (transparent gender congruent) in comparison to the latter (transparent gender incongruent), yet none of the three experiments showed this, not even marginally[38]. It is hence more likely that the effect obtained in Experiment 2, in which opaque targets benefited from the presence of transparency congruent distractors in comparison to the presence of transparency incongruent distractors, is directly related to the routes of gender retrieval and so can be considered a reflection of the situation we have described for Experiment 1. In any case, note that this explanation is wholly tentative. In fact, we could not replicate this effect of Experiment 2 in Experiment 3 and hence (a), the effect might be a fluke, or (b), the effect might simply be hard to obtain – perhaps due to the complexity of the PWI paradigm. Additionally, Experiment 3 might have completely failed to obtain any effect due to the

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

[38] In this sense, there is evidence that the classical ortho-phonological facilitation effect in the PWI paradigm occurs based on the final phonemes of the stimuli pairs. Yet this evidence is mainly obtained for the sharing of the final syllable or last two letters/phonemes when stress patterns are controlled (e.g., Ayora et al., 2011; Bürki et al., 2016; Meyer, 1996; Meyer & Schriefers, 1991; Wilshire et al., 2016). It is far from clear whether the sharing of only the final letter with mixed stressing patterns between targets and distractors could produce any kind of visible facilitation effect.

inclusion of animate targets within the stimuli list, similar to what we observed in the second experiment of our first study with 25% of animate targets. Here, we had 26% of animate targets and 9% of animate distractors. As we have discussed, in theory, animate stimuli randomly included within the materials might somehow influence the outcome of the experiment – yet, (1) Cubelli et al. (2005), Paolieri, Lotto, et al. (2010), and Paolieri et al. (2011) used similar percentages of animates and systematically found the GIE; (2) our GCE in Study 1, on which we based the empirical confirmation of our rationale regarding animacy, does not seem to be reliable.

Critically, there are two considerations that we believe must be made regarding our results before closing this section, one concerning the nature of grammatical gender selection and the other having to do with the GIE. Regarding the former issue, although we do find clear supporting evidence for the DSIH, we do not think that we are in position to view our results as supporting evidence for the notion of gender as an automatically retrieved characteristic represented by nodes that do not compete. The reasons for this have been set out previously in this thesis. In sum, the PWI paradigm might be able to reflect the competition happening between determiners, but not between gender nodes. We will address this debate in more detail in the General Discussion.

Finally, we still face the same enigma that we encountered in the previous study: the reasons behind our systematic failure to observe a GIE that is consistently obtained with Spanish and Italian participants in the experiments of Cubelli et al. (2005), Paolieri, Lotto, et al. (2010), and Paolieri et al. (2011). We would like to highlight two facts regarding this. On the one hand, our results are in line with those of Finocchiaro et al. (2011), and thus they constitute evidence against the GIE. Theoretically, this would be a relief: as stated multiple times in the current thesis, it is difficult to explain the GIE and fit it with evidence on gender priming and bilingualism, and the DSM does not offer a theoretically adequate approach to do so. Yet our meta-analysis confirms that the GIE is a significant effect, be it very small and heterogeneous, and hence we cannot disregard the possibility that our results simply fall into the realm of the null results that are expected when an effect is of such a small size and high in heterogeneity. Indeed, an update of our meta-analysis including our two studies (in the case of our first study, only the conditions with 0 and 25% of animate targets) reveals that the GIE is small but still significant (with k = 37, g = -0.108, SE = 0.049, CI [-0.212, -0.003], t(14.67) = 2.20, p = .044). Nevertheless, significance decreased. We think that further replication of these

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

experiments will provide us with the answer we seek regarding the reliability of this effect. On the other hand, an inspection of our data reveals a critical situation that had already been noted in our first study: the direction of the potential gender-based effect is extremely unstable, and there might be some type of individual differences determining whether what emerges is a GCE or a GIE. More specifically, in our first experiment, 42 of the participants showed an effect of GC, but in fact 34 of them showed a tendency for a GIE. In our second experiment, 43 of the participants tended to show a GCE, but 36 a GIE. Importantly, in the third experiment, only 14 participants tended to show a GCE, the other 35 tended to show a GIE (mean = 27ms). Although what we are assessing here are non-significant values, we do not think these can be ignored in such a convoluted area in which the overall results have reflected every possible situation: the same effect in opposite directions and null effects. The possible existence of individual differences underlying such mixed tendencies across participants will be assessed in more detail in the General Discussion.

6.4 Study III: Cross-linguistic GCE and the impact of proficiency

The assessment of native speakers of European Portuguese through the PWI paradigm reveals that the GCE obtained in our first study is indeed not robust and it might be a fluke. We have not been able to replicate the GIE either. Our results seem to support the DSIH. Yet although we have not observed gender-based effects in European Portuguese, according to the literature we should obtain a cross-linguistic GCE when assessing bilinguals of this language (L1) and another (L2) with either naming or forward translation tasks. The effect should be observed regardless of the language family, as well as of the transparency of the language, and the presence of agreement. We therefore conducted an experiment in which we tested bilinguals of European Portuguese and German through a translation task in both translation directions (European Portuguese to German and vice-versa) using bare nouns. Note that German is also a language for which genuine effects of gender in the PWI paradigm have not been obtained. Ultimately, our main concern is not only to replicate the cross-linguistic GCE for the first time in European Portuguese, thus finding support for the idea that gender nodes compete, but also to assess the role of proficiency. Proficiency has been shown to be critical according to empirical evidence on L2 processing, but the role it plays here is unclear most likely due to a poor assessment of this variable in the literature, as seen in our meta-analysis (c.f., Bultena et al., 2015; Lim & Christianson, 2007; Prior et al., 2011).

USC UNIVERSIDADE DE SANTIAGO

To make predictions regarding the influence of proficiency in gender retrieval in bilinguals, first, we assumed that gender is represented as nodes located at a specific level of lexical access, either the lemma or an independent syntactic network. We also assumed that gender nodes accumulate activation and compete for selection. To predict how grammatical gender representation and processing would change depending on proficiency, we followed the BIA-d model (Grainger et al., 2010) and the Multilink (Dijkstra et al., 2019). The BIA-d model offers an interesting view on the adaptations that our linguistic system experiences during the acquisition of an L2. More specifically, when the L2 increases its proficiency, it decreases its dependency on the L1, and hence decreases the ability of the L1 to create interference during L2 processing (Abutalebi & Green, 2007; Pivneva et al., 2012). Adding the tenets of the Multilink to our rationale, as the frequency of encounters with words increases, the stronger the links between their lemmas and their grammatical features. We hypothesize that this would also occur with grammatical gender. The stronger the link between L2 lemma and gender nodes, the easier the retrieval of gender and the less the interference of the L1 translation. Indeed, it is known that the higher the L2 proficiency of the bilingual, the better their performance during L2 production at different levels (Costa & Caramazza, 1999; Kroll et al., 2002; Pivneva et al., 2012). We hence believe that this rationale should apply to this field, and hence the cross-linguistic GCE should interact with a measure of proficiency if the latter is properly assessed with a highly valid and reliable measure. Note, however, that our sample consisted of early bilinguals, barely tested in the field (only – presumably – in Costa, Kovacic, Franck, et al., 2003; Manolescu & Jarema, 2015; Paolieri et al., 2020). It has often been said that if acquired early, bilinguals have better control of the influence of one language over another, yet age of acquisition is not determinant for this control. Other kinds of factors, such as daily exposure to each language or changes in the language environment, can modulate the interaction of both languages during language processing regardless of when the L2 was acquired (for a review, see Van Hell & Tanner, 2012; see also Bonfieni et al., 2019). This occurs largely because these factors directly influence the degree of proficiency of the speaker in each language. Therefore, whether or not we observe a cross-linguistic GCE should depend mainly on the degree of proficiency in each language, and being an early bilingual should not be an impediment to language interaction happening. Indeed, in Paolieri et al.'s (2020) the cross-linguistic GCE was clearly observed, even though participants were early bilinguals (but where one language was dominant over the other). We thus believe that modulations in proficiency of either of the two

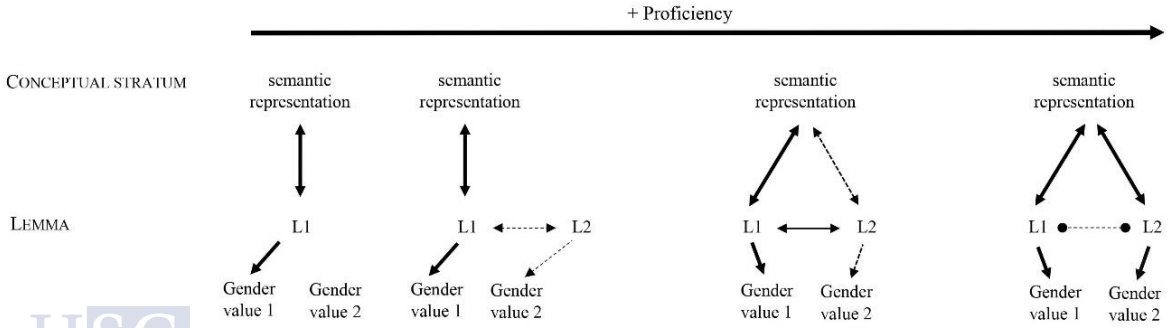
USC UNIVERSIDADE DE SANTIAGO

languages should produce changes in the strength of the links that each lexical entry has within each stratum (including the links connecting grammatical gender to lemmas). The changes in the strength create states of dependence or interference of one language over another, somehow regressing in the phases of acquisition, as suggested by the BIA-d model (for such phases, see our proposal in Figure 16). In this sense, note that either the L2 or the L1 could suffer changes in their representational state, depending on these fluctuations in proficiency, in line with previous evidence (e.g., Dussias & Sagarra, 2007; Guo et al., 2011; Linck et al., 2009; see also Morales et al., 2014).

In the present study, we recruited participants from Portuguese families that were born or lived during childhood in either Germany or Switzerland (i.e., German speaking territories), whether or not they stayed in these countries later in life. Consequently, they had different degrees of proficiency in each language, suffering certain imbalances between languages due to a reduced proficiency in either their L1 or their L2. As hypothesized, in some of these cases, the strength of the links between the gender nodes and the L1 lemmas had likely decreased, creating a situation of possible interference between languages, even though they were early bilinguals and highly proficient in the L2. We hence considered not only the proficiency of the L2 but also of the L1, using the Dialang for a standardized and highly valid and reliable measurement of the degree of proficiency in both languages (see Alderson, 2006). As an imbalance between languages seems to be key for interference to occur, we considered it by subtracting the Dialang score in one language from the score in the other.

Figure 16

How L2 gender representation develops during acquisition following the BIA-d model



Note. Instead of the lexeme, we opted to represent the lemma connected through unidirectional links to gender nodes. L1 = First language; L2 = Second language. Discontinuous lines represent weak connections. The thinner the line, the weaker the connection. In our predictions,

the representational state of the linguistic system may vary depending on the proficiency of one language or another. Figure based on Grainger et al. (2010).

6.4.1 Method

6.4.1.1 Participants

Forty-nine early bilinguals of European Portuguese and German (39 female; M_{age} = 38.25 years, *SD* = 9.46) were recruited voluntarily online through email and social media. All were from Portuguese families but born in Germany or that had emigrated during their childhood to Germany. They acquired German before the age of 10 (47 participants acquired German before the age of 7; M_{AoA} = 2.29 years, *SD* = 2.25). Ten of them reported moderate to high proficiency in another gendered language apart from German or Portuguese in the Language History Questionnaire (LHQ, Li et al., 2020). These were French, Spanish and Italian. All participants signed informed consent for experimentation with human subjects previously approved by the Ethics Council of the University of Minho (CEICSH 120/2020) through Google Forms.

6.4.1.2 Materials

We selected 180 European Portuguese inanimate nouns from the P-PAL database (Soares, Iriarte, et al., 2018) and translated them into German. They were selected by taking into account the gender value in both languages, so that we had the same number of stimuli in each of these 6 translation types: heterogeneric feminine-masculine ("abóbora" [F] in Portuguese, "Kurbis" in German [M], "pumpkin"), heterogeneric masculine-feminine ("journal" [M] in Portuguese, "Zeitung" in German [F], "newspaper"), homogeneric feminine ("cenoura" in Portuguese, "Karotte" in German, "carrot"), homogeneric masculine ("bosque" in Portuguese, "Wald" in German, "forest"), feminine-neuter ("perna" in Portuguese, "Bein" in German, "leg"), and masculine-neuter ("carro" in Portuguese, "Auto" in German, "car"). We avoided European Portuguese nouns with more than one German translation that had similar frequencies of use according to SUBTLEX-DE (Brysbaert et al., 2011). Besides, we avoided nouns in German that could also be verbs (e.g., "Leben" [life/to live]), nouns with high positive or negative affective valence related to death or sexuality (e.g., corpse, death, penis, etc.), German nouns that were cognates in English (e.g., "Butter"), and nouns that had more than one possible translation and when these diverged in gender within that language (e.g., "Miete" [rent,

USC UNIVERSIDADE DE SANTIAGO

F] in German translates to "Renda" [F] or "Aluguer" [M] in Portuguese). In terms of ortho-phonological gender transparency, we did not select any irregular nouns in Portuguese, and included 144 transparent nouns and 36 opaque nouns distributed evenly across the six translation types (24 transparent nouns and 6 opaque nouns per type). We created two different blocks as a function of translation direction (European Portuguese to German and vice-versa), each composed of 90 stimuli, so that the presentation of both blocks was counterbalanced across participants. All Portuguese nouns were controlled through one-way ANOVAs across translation types (6 conditions) for per million and logarithmic frequency, number of phonological and orthographic neighbours, number of letters, and mean logarithmic bigram frequency, taken from P-PAL (all *p*s > .247, Soares, Iriarte, et al., 2018), logarithmic frequency, taken from SUBTLEX-PT (p = .544, Soares et al., 2015), and subjective frequency, concreteness, and imageability, taken from the Minho Word Pool (all *p*s > 525, Soares et al., 2017). See mean values in Table 30.

Table 30

Means and standard errors of the controlled variables for Portuguese nouns across translation types

	HET F-M	HET M-F	HOM F	HOM M	N-F	N-M
FpM	23.71	71.62	39.01	17.78	30.63	81.30
	(6.83)	(31.49)	(14.57)	(4.20)	(13.14)	(40.79)
LogF	1.02	1.13	.96	.99	.98	1.18
	(.10)	(.14)	(.13)	(.09)	(.11)	(.14)
PN	4.97	5.03	2.87	4.27	3.93	4.07
	(.77)	(1.09)	(.63)	(1.09)	(.81)	(1.07)
ON	6.03	4.00	2.70	2.87	4.20	4.90
	(1.05)	(.86)	(.60)	(.80)	(1.08)	(1.23)
NL	6.03	6.00	5.93	6.40	6.20	6.33
	(.36)	(.39)	(.26)	(.29)	(.32)	(.36)
MLBF	2.69	2.60	2.47	2.46	2.58	2.46
	(.08)	(.06)	(.09)	(.06)	(.07)	(.09)
LogF-S	3.09	3.07	2.89	3.04	3.05	3.26
	(.12)	(.14)	(.14)	(.09)	(.15)	(.14)
SubjF	4.82	5.13	5.08	4.69	4.92	4.89
	(.20)	(.23)	(.19)	(.20)	(.24)	(.26)
Concret	5.94	6.15	6.18	6.02	6.13	6.02
	(.21)	(.19)	(.17)	(.23)	(.17)	(.16)
Imag	5.78	5.87	5.85	5.55	5.78	5.63

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

	(.13)	(.12)	(.08)	(.17)	(.14)	(.16)

Note. FpM = Frequency per million; LogF = Logarithmic Frequency; PN = Phonological number of neighbours; ON = Orthographic number of neighbours; NL = Number of letters; MLBF = Mean logarithmic bigram frequency. All obtained from the P-PAL database (Soares, Iriarte, et al., 2018). LogF-S = Logarithmic Frequency. Obtained from the SubtLex-PT database (Soares et al., 2015). SubjF = Subjective Frequency; Concret = Concreteness; Imag = Imageability. All obtained from the Minho Word Pool database (Soares et al., 2017).

German nouns were controlled across the six translation types through one-way ANOVAs for absolute logarithmic frequency, number of letters, logarithmic number of numbers based on the Levenshtein distance, initial logarithmic bigram frequency (normalized), and familiarity, as taken from dlexDB (all *p*s > .091; Heister et al., 2011), and logarithmic frequency as taken from SUBTLEX-DE (Brysbaert et al., 2011; $p = .29$). See Table 31 for means and standard errors. Translation pairs across translation types were controlled for equivalent measures, namely number of letters, and subjective frequency/familiarity (*p*s > 563). Logarithmic frequency (SUBTLEX-PT and SUBTLEX-DE) was considered later, and although it showed significant differences, these differences were not between the conditions that subsequently showed significant results (i.e., heterogeneric translations; *p*s >.143). The translations were also controlled for orthographic overlap using the NIM database (Guasch et al., 2013) and phonological overlap using the PHOR-in-One database (Costa et al., 2021; all *p*s > .174). See Table 32 for means and standard errors of overlap measures.

Conditions regarding the stimuli were created taking into consideration the factors of Gender Congruency (gender congruent, gender incongruent), Target Gender (masculine, feminine), and Target Language (German, Portuguese). Note that target gender and target language refer to the gender of the noun to be produced, and the language to be produced, respectively. Importantly, the factor of gender congruency included the four types of heterogeneric (incongruent) and homogeneric (congruent) nouns. This allowed us to make a direct comparison between the gender systems of European Portuguese and German. Neuter gender in German would constitute a third category, in which rather than gender incongruency there is a situation in which one gender value does not exist in the other language, and this may change the representation of that value and the interaction between languages when it comes to its retrieval. The scope of our study is to replicate the cross-linguistic GCE and test the activation of gender nodes and the processes of competition that may arise between them

depending on proficiency. These aims can be fulfilled with heterogeneric and homogeneric nouns, avoiding a greater degree of complexity in our design. Nevertheless, we included neuter nouns among our stimuli from the beginning, to avoid artificial contexts that might in some way influence the performance of our participants.

Table 31

Means and standard errors of the controlled variables for German nouns across translation types

	HET F-M	HET M-F	HOM F	HOM M	N-F	N-M
LogF	3.27	3.63	3.28	3.47	3.41	3.63
	(.16)	(.23)	(.31)	(.14)	(.24)	(.22)
Neigh-L	4.87	5.19	4.58	4.75	5.32	5.29
	(.41)	(.48)	(.38)	(.44)	(.62)	(.53)
NL	5.83	6.13	5.80	5.50	6.63	6.10
	(.35)	(.46)	(.24)	(.36)	(.41)	(.46)
ILBF	4.64	4.66	4.21	4.30	5.39	5.19
	(.29)	(.33)	(.30)	(.30)	(.41)	(.39)
Fam	5.03	4.84	4.38	4.43	5.30	5.23
	(.31)	(.37)	(.34)	(.29)	(.46)	(.40)
LogF-S	2.52	2.56	2.29	2.53	2.50	2.73
	(.11)	(.13)	(.13)	(.09)	(.15)	(.14)

Note. LogF = absolute logarithmic frequency; Neigh-L= logarithmic number of numbers based on the Levenshtein distance; NL = Number of letters; ILBF = Initial logarithmic bigram frequency; Fam = Familiarity. All obtained from the dlexDB database (Heister et al., 2011). LogF-S = Logarithmic Frequency obtained from the SubtLex-DE database (Brysbaert et al., 2011).

Table 32

Means and standard errors of the overlap measures

	HET F-M	HET M-F	HOM F	HOM M	N-F	N-M
Ort-Over	.15	.20	.23	.18	.13	.24
	(.03)	(.04)	(.04)	(.03)	(.03)	(.03)
Phon-Over	.39	.46	.47	.41	.43	.43
	(.03)	(.03)	(.04)	(.03)	(.03)	(.03)

Note. Ort-Over = Orthographic overlap. Obtained from the NIM database (Guasch et al., 2013). Phon-Over = Phonological overlap. Obtained from the PHOR-in-One database (Costa et al., 2021).

6.4.1.3 Procedure

The experiment was conducted online due to the public international health emergency caused by the COVID-19 pandemic. We followed Burke & James' (2006) recommendations for online research and data collection.

Participants started by filling out the LHQ (Li et al., 2019) using Google Forms. This questionnaire allowed us to explore their linguistic background and check their knowledge of any other gendered language. Afterwards, links were sent for the Portuguese and German versions of the Dialang, a standardized lexical test to objectively assess the proficiency in each of these languages while guaranteeing a high degree of validity and reliability.

Participants then received a link that opened the experiment in a browser. The experiment was programmed using the JavaScript library jsPsych (de Leeuw, 2015). Each link included the two blocks of nouns (in total, 190 nouns), but the order of presentation and the language to be produced (i.e., the target language) was counterbalanced, so that we had four different links depending on these two factors. After clicking on the link, the participant first read the instructions regarding the procedure for the whole experiment. Instructions appeared in the target language of the first block. Then, a familiarization phase started in which the 90 translation pairs of the first block were presented sequentially one by one. Participants controlled the presentation of the stimuli using the spacebar. The aim of the familiarization phase was to decrease mistranslations and non-responses. After that, the participant tested their microphone, following instructions on the screen, and once checked, instructions appeared for the translation of the first block. The participants were instructed to translate each noun into the target language as fast as possible avoiding mistakes and trying to speak loudly and clearly into the microphone. Upon starting, participants went through a session of eight training items (different from the experimental items), then the experimental trials started. Each experimental trial had the following structure: a fixation point (+) at the centre of the computer screen, for 500 ms; the target picture with the superimposed distractor for 2,000 ms or until response; a blank space for 500 ms as an inter-trial interval. Trials were presented randomly per participant. For the second block, instructions appeared in the other language (which would be the new target language). The familiarization phase and translation task occurred again with the remaining 90 nouns. Responses were recorded and saved in a private directory on the University of Rovira i Virgili server. Response times were calculated offline from the presentation of the

USC UNIVERSIDADE DE SANTIAGO

noun to be translated to the onset of the translation response using the PRAAT software (Boersma & Weenink, 2018).

The session for each block lasted approximately 20 min.

6.4.2 Results

We removed the RTs of incorrect responses, RTs above 2,000 ms, and RTs that exceeded 2.5 *SD* of each participant's mean. We also removed the data from seven participants that committed more than 40% of incorrect responses (the final sample was composed of 42 participants). RTs were analysed using linear mixed-effect models (e.g., Baayen, 2008; Baayen et al., 2008). To this end, we used the lme4 package of R (Bates et al., 2015). We created several models in which we introduced the following variables (and some of their interactions) as fixed effects: Gender Congruency (GC, GI), Type of Gender Congruency (heterogeneric, homogeneric), Target Language (language that was produced), Target Gender (gender of the noun that was produced), Age of Acquisition (AoA) of the German Language, Contact with German Language (years, in logarithm scale), and Absolute Difference in Proficiency between the Portuguese and German Languages (through Dialang scores, "difference in proficiency"). In addition, random intercepts for participants and words were included in each model, as well as by-participant random slopes, following a maximal random effects structure (Barr et al., 2013). However, in most cases this maximal random effects structure did not converge or produced a singular fit (i.e., some random effects explained a near-zero variance), and we had to remove some random slopes from the model to achieve convergence. Random slopes were kept in the model only if they increased its fit significantly. We also included the results of the *t*-test analyses of the coefficient estimates for each fixed effect. To this end, we used Satterthwaite approximations to the degrees of freedom of the denominator (*p*-values were estimated by the LMerTest package, Kuznetsova et al., 2017). Mean response latencies are presented in Table 33; estimated fixed effects for the final model are presented in Table 34.

Table 33

Mean RTs and SEs of gender congruency according to target language and target gender

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Target language	Gender congruency	Mean	SE
GER	GC	1104	8.2
GER	GI	1115	8.4
PT	GC	1094	8.1
PT	GI	1087	7.9

Table 34

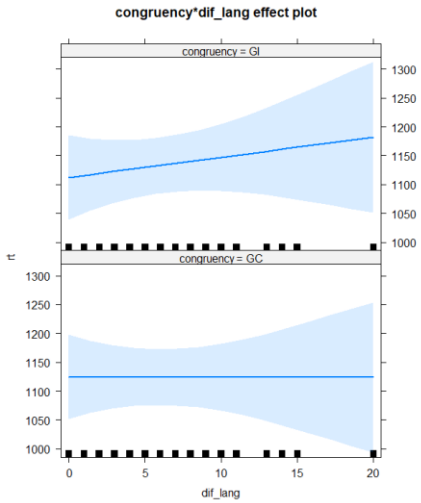
Estimated fixed effects for the final model

	Estimate	SE	t	*p*
(Intercept)	1111.44	73.85	15.05	<0.001
gender congruency GI	-13.55	24.13	-0.56	0.574
target_gender M	8.82	11.05	0.80	0.425
target_language PT	**-17.13**	**8.32**	**-2.06**	**0.039**
AoA_German	-9.76	7.89	-1.24	0.216
dif_prof	0.00	4.43	0.00	1.000
contact_German	12.25	22.48	0.55	0.586
gender congruency GI:dif_prof	**3.62**	**1.47**	**2.47**	**0.014**

The results revealed no main effect of cross-linguistic GCE, nor of any other factor included in the analysis (all *p*s > .05) except for Target Language, as participants were faster when producing nouns in Portuguese than when producing nouns in German, estimate = -17.13, *SE* = 8.32, *t* = -2.06, *p* = .039. Importantly, we obtained a significant interaction between the factors Gender Congruency and Difference in Proficiency $\chi^2(1) = 6.08$, $p = .014$. When participants were balanced (Difference in Proficiency ≃ 0), no cross-linguistic GCE was observed; however, for those participants that were more proficient in one language than the other, a cross-linguistic GCE was observed. This cross-linguistic GCE increased in parallel with the difference in proficiency between languages (i.e., the higher the imbalance, the higher the cross-linguistic GCE, see Figure 17). Moreover, the effect is related to incongruent pairs: congruent pairs do not add facilitation, but incongruent pairs generate interference.

Figure 17

Interaction between gender congruency and difference of proficiency between languages



USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Note. Reaction times in milliseconds. GC = Gender congruent; GI = Gender Incongruent; dif_lang = Difference in proficiency.

6.4.3 Discussion

The results confirmed the existence of a cross-linguistic GCE with bare nouns in a translation task in the two translation directions with bilinguals of European Portuguese and German. Hence, they support the ideas behind the GSIH, in the sense that genuine effects of gender competition are obtained. Moreover, these gender effects are observed in two languages in which they had not been observed before in the PWI paradigm (except for the GCE in our first study, which lacked robustness). Therefore, our results speak in favour of the idea that gender nodes gather activation, compete for selection, and are shared across languages. They hence go against the architecture of lexical access proposed by the DSM and are in line with the architecture proposed by WEAVER++ (except for the requirement of agreement for gender selection). Importantly, the cross-linguistic GCE occurred regardless of the target language, that is, it happened from German to Portuguese and vice-versa. Since the effect was obtained in both directions, it seems that the imbalance in terms of proficiency in any language creates interference. Even though differences in proficiency exist for our participants, and thus they have variations on the strength of their lemma-to-gender links between languages, they are highly proficient in both languages. Hence, the links for the less proficient language are still strong enough to generate interference during the processing of the more proficient language. This may not be the case if the speakers have very low proficiency in one of the languages, thus having very weak links, as in the first phases proposed by the BIA-d model (see Figure 16).

Importantly, note that AoA of the L2 German (from 0 to 10) was not significant, thus confirming our idea that the "when" is not the determining factor for whether interaction between languages occurred during gender selection, but that variations in proficiency of both languages is. In this sense, our results are also in line with Paolieri et al. (2020), who found a clearly cross-linguistic GCE in language comprehension for early bilinguals who show differences in dominance in their two languages.

USC DE SANTIAGO DE COMPOSTELA

Finally, note that the cross-linguistic effect was of interference: unbalanced bilinguals did not experience facilitation when translations were homogeneric; they experienced increased interference for gender retrieval when translations were heterogeneric. This is interesting in the sense that, if gender nodes were not competing, and were only accumulating activation to be

automatically selected, as Caramazza et al. (2001) argue, they should not interact with each other and an effect of GCE should be observed only in terms of facilitation. This is so because even if they do not compete, activation coming from different sources should add to each other and mean that reaching the threshold for selection is faster. Yet in fact we observe the opposite here, which strongly suggests competition.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

7 GENERAL DISCUSSION

The literature on grammatical gender processing during language production has reported intricate findings. Studies on native language processing have relied on a complex paradigm that brings together language production and word recognition: the PWI paradigm; studies on L2 production, however, have relied on simpler tasks such as naming and forward translation tasks. Both fields contribute to the heated debate as to the mechanisms underlying the selection of grammatical gender through the assessment of gender congruency effects. Studies on native language production introduced the discussion regarding GSIH vs. DSIH: whereas the former interprets the outcomes of the PWI paradigm as reflections of the competition between abstract gender values, the latter sees them as competition between determiners. The assessment of both hypotheses has been a means of exploring the core mechanisms underlying gender selection. Authors supporting the GSIH conceive gender values as nodes that compete for selection, but the authors defending the DSIH conceive of gender as an automatically selected characteristic not subject to competition. Adopting one view or another is not an easy task since results are unclear within native speakers and mixed when considering the type of population ("monolingual" vs. bilingual).

A meta-analytic review of the literature confirms that the so-called classical "GCE" found in Germanic and Slavic languages is actually small but significant and seems to support the determiner congruency effect and hence the DSIH. This because it is only obtainable when participants name the pictures using noun phrases with freestanding elements of agreement that vary across gender values. This conclusion could benefit from a meta-analytic confirmation of the selection of gender in Dutch and German during the production of definite articles for the plural and diminutive forms. The resulting confirmation here would provide firm support for the idea that the absence of a "GCE" when the elements of agreement do not vary across gender values (i.e., when using either the plural or the diminutive form, "die" [German]/"de" [Dutch] and "het" [Dutch], respectively) is due to the presence of only one agreement form for determiners across genders, and not due to gender not being selected. Unfortunately, it was not

USC
DE SANTIAGO
DE COMPOSTELA

possible to carry out this type of analysis, as there are currently insufficient studies assessing this question through the singular-plural paradigm to obtain significant and interpretable results. It is worth noting, however, that certain studies with more fine-grained techniques, namely fMRI, found competition between gender values to be the source of competition on the "GCE" (Heim, Eickhoff, et al., 2009; Heim, Friederici, et al., 2009). The authors interpreted this as the fMRI possibly reflecting gender as a first source of competition, even if what is reflected on the RTs afterwards is competition between determiners.

Studies with Romance languages paint an even more difficult picture on the matter. When participants name the pictures using noun phrases, the "GCE" is consistently absent in languages such as Italian and Spanish. In these, a one-to-multiple gender-to-determiner mapping opposes the one-to-one mapping of other languages such as German and Dutch. The selection of one determiner form or another within the same gender value depends on the stressing patterns and the first letters of the following word, which can be either a noun or an adjective. Thus, null effects are a consequence of the additional processes of syllabification and word organization that take place to make possible the selection of the proper determiner within a specific gender value (Late Selection Hypothesis, Miozzo & Caramazza, 1999). These processes mask – in terms of RTs – the competition that occurs between determiners of different gender, rending the "GCE" experimentally invisible.[39] Among Romance languages, an exception is French, in which the classical "GCE" has been observed when either varying the SOA (Foucart et al., 2010; but for null results see Bürki et al., 2020) or taking the reading times of the participants into consideration (Bürki et al., 2020). French has one definite article per value, and the same reduced form "l" for both values when the noun starts with a vowel. Additional processes take place, but since the extra determiner form is not related to gender (it is the same form for both values), the selection of the determiner supposedly requires simple phonological transformations and occurs earlier than in Spanish and Italian, making the "GCE" observable under certain specific conditions that outperform the "late" selection of the determiner. Our studies with European Portuguese, a transparent Romance language with a one-to-one mapping, confirm the Late Selection Hypothesis, since a "GCE" was observed in our second study with noun phrases and the typical SOA of 0 (simultaneous presentation of target and distractor). Our study thus supports the idea that the differences in the results between both

[39] We could say that the encoding of the noun form is also one of the processes delaying gender selection, but the results in Bürki et al.'s (2016) EEG experiment suggest that noun's form is always processed before determiners, regardless of the possibility for an early determiner selection (in line with the IN model).

groups of languages is due to gender-form mapping and not due to other characteristics associated with being Romance or transparent. However, our study also confirmed that the "GCE" in transparent languages might not be as simple as it is in opaque languages. Our "GCE" was only present for opaque nouns, especially when paired with transparent distractors. The additional form-based route of transparent nouns seems to make them more prone to interfere with the selection of gender of another noun, and also less prone to receive interference from a distractor noun. Indeed, it has been shown that transparent nouns consume more cognitive resources during lexical access than opaque nouns (Caffarra & Barber, 2015; Caffarra et al., 2014, 2015). Conversely, opaque nouns are not as safe from interference since they only rely on the lexical route. The different amounts of activation coming from both routes create different patterns of activation that seem to cascade down to the determiner forms available for selection, generating interaction between the determiner congruency effect and the transparency of targets and distractors. We hence obtained supporting evidence that the routes of gender retrieval used during language comprehension also have a role during language production.

When speakers of Italian and Spanish name the pictures with bare nouns, a very small but significant GIE emerges. Yet the effect has a high degree of variability, and this might explain why many experiments have failed to observe it in these languages, as well as in other Romance languages, namely French (Finocchiaro et al., 2011) and European Portuguese (our own studies). Another factor that perhaps explains this variability is the existence of individual differences in which tendencies for both sides of the effect (GCE and GIE) are observable and may be cancelling each other. In this sense, we did obtain a significant GCE (Experiment 1 of our first study), but according to the Bayesian analyses conducted afterwards it was not robust enough, and we were unable to replicate it.

In sum, the picture for Romance languages is more chaotic than that for Germanic and Slavic languages. Studies with European Portuguese and French seem a priori to support the DSIH, this because effects are not obtained with bare nouns, but they are obtained with noun phrases when it is experimentally possible to do so. Studies with Italian and Spanish participants, however, mainly observe a GIE but support neither the DSIH nor the GSIH, since what seems to be competing and causing the GIE are nouns of the same gender, rather than gender nodes.

USC UNIVERSIDADE DE SANTIAGO

In theoretical terms, those models that have addressed gender in their tenets either fail to predict the results or provide unsatisfactory explanations of those results. Both the WEAVER++

and IN models conceive of gender values as nodes that accumulate activation and compete for selection. WEAVER++ specifies the concepts of "critical difference" and "threshold of selection", arguing that gender selection only occurs when other words that require agreement are present, and affirms that gender retrieval takes place (or at least starts) before word-form processing. The authors of the IN model are not as specific in their formulation but conceive of competition and inhibition between gender nodes and affirm that gender selection occurs after lexeme encoding has taken place (but word-form processing, understood as the processing of the ortho-phonological segments of the noun, would take place in parallel to gender selection). They do not address the role of agreement but make possible the skipping of grammatical encoding if necessary – hence the idea of WEAVER++ regarding the obligatory nature of agreement could fit the architecture of the IN model (i.e., if agreement is absent, gender encoding is skipped). Both models seem to be imprecise in their approaches when trying to explain the results described above. In Germanic and Slavic languages, gender appears not to compete, and agreement is not a necessary condition for gender to be selected (see the results in the literature with the singular-plural paradigm). This has led certain authors to make a ground-breaking claim: gender is a feature that is selected automatically without competition occurring between values (Caramazza et al., 2001). In Romance languages, the effect that continues to be significant goes in the opposite direction of what we would expect, because stimuli of the same gender generate effects of competition rather than of facilitation. Hence, what seems to be competing for selection are nouns themselves rather than gender nodes. This is not supported by either the WEAVER++ or the IN model. Moreover, the DSM (Cubelli et al., 2005), a more recent proposal focusing on gender itself, sought to explain the mixed results in different languages in terms of the variable of gender transparency, but it relies on a series of theoretically imprecise rationales. Of note, this model puts forward the idea that nouns of the same gender are always activated together and hence always compete for selection. If this were so simple, we do not think that the GIE should be so hard to obtain. In fact, our meta-analysis revealed that the effect is actually very small and highly heterogeneous.

When bilingualism is brought into the picture, our understanding of the process by which gender values are retrieved becomes more confusing, because results go against what is seen with native speakers. Yet if taken alone (putting aside the PWI paradigm), they are actually quite simple. A small cross-linguistic GCE (but not as smaller as the GIE) that depends on proficiency is consistently observed, regardless of the language family and the presence of

USC UNIVERSIDADE DE SANTIAGO

agreement; hence a genuine gender effect. Support is thus given to the ideas behind the GSIH and to an integrative view of the bilingual gender system, nodes being shared across languages and accumulating activation from lexical entries of both languages. Results oppose the idea that gender is automatically retrieved and that an agreement context is compulsory for gender retrieval to happen. Critically, the cross-linguistic GCE completely contradicts the GIE and the views of the DSM. If the gender system of bilinguals is integrated, and hence all nouns are connected to the same gender nodes – understood as semantic features by the DSM –, translations of the same gender are connected to the same gender node, which would increase their similarity (like cognateness would) and hence they should compete for selection harder than when they are of different genders. Yet they do not do this. Heterogeneric translations generate more interference during lexical selection than homogeneric ones, and therefore it seems that it is a matter of gender itself: gender values are competing for selection rather than nouns within the same shared gender system, and hence when translations are heterogeneric, what the WEAVER++ model would call the "critical difference" is harder to reach and the cross-linguistic GCE arises.

The challenge now is to try to find answers for the clash of results between native language and L2 production, and for the unclear results of the studies with native speakers. We will base our arguments on the picture we have just drawn and will provide a series of tentative hypotheses of the mechanisms underlying gender selection.

7.1 HYPOTHESES FOR THE MECHANISMS UNDERLYING GENDER SELECTION

We will propose two different hypotheses based on two different assumptions in order to explain the underlying mechanisms of gender selection. We will start by assuming that gender nodes do not compete for selection; we will then assume that they do compete for selection.

7.1.1 Gender nodes do not compete for selection

If we accept the claims of Caramazza and colleagues (2001; see also Schiller & Caramazza, 2003, 2006), genuine effects of gender with bare nouns within the PWI paradigm in Germanic, Slavic, and Romance languages are absent because gender nodes do not compete. Competition between determiners that vary across gender values is the only robust effect we can obtain. Following WEAVER++, we could argue that the absence of competitive effects when participants produce elements that do not vary across gender values, such as the plural definite

USC UNIVERSIDADE DE SANTIAGO

articles in German and Dutch, is due to agreement not being required to establish the form of these elements. Since agreement is not required, gender is not retrieved and hence, gender nodes do not compete for selection. Yet this reasoning does not hold because it has been shown that gender is being selected in these cases (i.e., with plural and diminutive forms in the singular-plural paradigm). Hence, the "GCE" that can be obtained within the PWI task is indeed exclusively a determiner congruency effect.

Regarding the effect of gender incongruency obtained for Italian and Spanish, in order to explain this we would have to assume that gender is a defining feature of nouns located at a shared level of syntax and semantics, and hence is not represented as gender nodes but as a property associated with each noun. To begin with, this means that gender values do not compete for selection, in line with our main assumption. Yet the DSM states that there is activation and selection of lemmas (which this model sees as representations of nouns that include all their conceptual and syntactic information), but never refers to selection by activation of gender itself. It is hence unclear how gender is selected. On the other hand, we have already addressed the flaws of this model when considering all the other assumptions that would have to be done to understand gender as a semantic feature. One way of solving this is to assume that gender is represented by nodes that are located in an independent syntactic stratum and that there are bidirectional links between nouns and gender nodes. In this way, gender nodes would function by collecting activation through different routes, and targets and distractors of the same gender would reinforce the activation of each other through the activation coming backwards from the same gender node and would hence generate the GIE. Yet this idea has two problems in light of the previous literature. In the first place, it does not match previous evidence on priming: if gender nodes had bidirectional connections, priming should be of inhibition between nouns of the same gender and not of facilitation (as seen in Heim, Eickhoff, et al., 2009, and Vigliocco et al., 2002). The second problem concerns the results on bilingualism.

Assuming either the idea of gender as a property not represented by nodes, the idea of bidirectional connections for gender nodes, or the non-competitive nature of gender values makes impossible for us to explain the cross-linguistic GCE. If we assume that gender is a semantic-syntactic property of nouns that contributes to the degree of similarity by which lemmas compete with each other, lemmas of different languages should see their similarity increased on the basis of gender, and a cross-linguistic GIE should be observed, which is not

USC UNIVERSIDADE DE SANTIAGO

the case. Conversely, if gender were a characteristic represented as nodes located in a specific syntactic network but had bidirectional links, translations of the same gender would reinforce each other, creating interference and generating the cross-linguistic GIE. This is also not the case. Additionally, if gender nodes did not compete for selection, we should not observe the cross-linguistic GCE, which is a genuine effect of competition between different gender values involving lexical entries of different languages. In sum, the only way of explaining the cross-linguistic GCE is assuming that gender values are represented by cross-linguistic shared nodes that compete for selection.

To sum up, assuming that gender nodes do not compete for selection offers a sufficient explanation of the results with Germanic and Slavic speakers, as well as French and Portuguese speakers, but leaves us without a thorough explanation for a (1) somewhat controversial GIE and (2) a whole set of consistent and theoretically plausible results with bilingual speakers. We will address the matter of the GIE in more detail in the next section, but regarding bilinguals, if we still want to maintain the assumption of the non-competitive nature of grammatical gender selection, we believe the only way to do so is to adopt a mixed view. We can assume that for "monolinguals" gender nodes do not compete for selection, but for bilinguals, they do. We are thus faced with a situation in which the core mechanisms of gender selection change depending on the number of languages we speak, which in our view might constitute an over-complicated system. In any case, one possibility is that entries of the same language activate gender nodes but do not make them compete, whereas entries of different languages activate these gender nodes and make them compete. Taking into consideration the BIA-d model (Grainger et al., 2010), since the entries belonging to different languages are actually equivalent translations in all the studies we have assessed, the L2 entry is created through the information already known for the L1 counterpart. Then, it slowly forms a fully specified independent representation that eventually stops relying on the information contained in the L1 entry. Hence, regarding gender, competition can be especially high when the gender node for the L1 entry is different from that of the L2 entry. More specifically, these two representations (L1 and L2) require the activation of different gender nodes, and this may create interference (competition for selection). Note that the Multilink model (Dijkstra et al., 2019) also offers a nicely complementary explanation on how the strength on the links between lemmas and other features change during language acquisition. If the strength of the links of the L2 entries with their gender nodes is still weak, the interference of the well-stablished links of the L1 entries will create more interference. The

USC UNIVERSIDADE DE SANTIAGO

interpretations based on the BIA-d and the Multilink model suggest that the proficiency of the bilinguals is hence determining whether gender nodes compete. Once a speaker starts to learn a L2, new entries join their lexicon as equivalent translations of already-existing entries, and gender selection, at least for the L2, transforms into a competitive process. As proficiency increases, the strength of the links between L2 entries and gender nodes would also increase, and independent representations for the L2 would form completely, which would consequently make gender nodes stop competing. Gender would hence start to be selected automatically for the L2. This is in line with the literature on bilingualism that finds that the lower the proficiency of the L2, the less automatic the production of the L2 is and the more it relies on controlling mechanisms to stop the interference coming from the L1 (note that this would apply in the reverse direction if the L1 is the less dominant language, as found in our third study; Pivneva et al., 2012; Segalowitz, 2010; Segalowitz & Hulstijn, 2005). Indeed, the cross-linguistic GCE we observed was an effect of interference coming from heterogeneric translations (those incongruent in gender). It is further possible that acquiring a L2 changes the core mechanisms of gender selection for all the languages of the bilingual (i.e., not only would gender be processed differently for the L2). This would mean that the cross-linguistic GCE should be obtained in both directions (L1 to the L2 and L2 to the L1). It is only when a high level of proficiency has been achieved in the L2, and the situation between both languages is balanced, that independent representations and strong links for each language are achieved and gender nodes stop competing. This idea is in line with our study and suggests that indeed, the core mechanisms of gender selection change for both languages (as the cross-linguistic GCE we found occurred regardless of which language was the target language, and only with a balanced situation competition stopped occurring).

Therefore, assuming a view of gender selection as a non-competitive process forces us to turn to a mixed view, one in which competition between gender nodes depends on the number of languages that the speaker knows, the state of acquisition of each of these, and his/her level of proficiency. Although this view allows us to explain our results quite well, some notable problems need to be taken into account. The first concerns the fact that this hypothesis violates the principles of Occam's razor in assuming changes in the core mechanisms of gender selection depending on the number of spoken (gendered) languages. The other problems are related to the existing evidence itself, namely (a) that the absence of evidence is the main pillar of this assumption despite the fact that it cannot be taken as evidence for the absence of effects (Vasisth

USC UNIVERSIDADE DE SANTIAGO

& Gelman, 2021), (b) the low number of studies with Germanic and Slavic languages supporting the idea that there is an absence of GCE, and (c) the existence of studies with more fine-grained techniques showing genuine effects of gender competition in monolingual populations (Heim, Eickhoff, et al., 2009; Heim, Friederici, et al., 2009). We will address these problems in more detail in the next sections.

7.1.2 Gender nodes compete for selection

By assuming that gender nodes compete for selection, there are two variables that must be highlighted to disentangle the results: the type of task, and the number of spoken languages (monolingual vs. bilingual speakers). We will first focus on the differences between tasks, and then on the contrast between monolinguals and bilinguals. The possible role of individual differences will also be considered.

Regarding the type of task used to study gender retrieval, the PWI paradigm has been shown to be less sensitive than a simpler paradigm, such as the singular-plural paradigm, to detect small effects of competition. Although the number of studies is not enough to establish robust conclusions, the evidence suggests that competition between bound morphemes is easily observed through the singular-plural paradigm. Besides, the singular-plural paradigm seems to be very effective in reflecting the patterns of competition between determiners of different number and gender during the production of a noun, as expected by the Primed Unitized Activation Hypothesis (Alario & Caramazza, 2002). These patterns seem to be harder to capture with the PWI paradigm, in which this type of competition should arise from not only the target noun itself but also from the distractor noun. A greater determiner congruency effect for the plural should be obtained when target and distractor have the gender value for which their singular and plural determiner forms are the same (common in Dutch, feminine in German). The scant literature here reveals that this type of effect is mostly absent (see all experiments of Schiller & Caramazza, 2003; but for the finding of such an effect, see Schiller & Caramazza, 2006). Similarly, it has been shown that in certain Romance languages, in which the mapping between gender values and determiners is more complex, gender-based effects (such as the GIE) and also medium-sized determiner effects (the "GCE") are not reflected in the RTs. This because these effects are masked by the processes of syllabification and word organization that take place after the distractor noun has been accessed. As a consequence, the RTs of the PWI

USC UNIVERSIDADE DE SANTIAGO

paradigm do not show these previous processes of competition between either gender values, noun lemmas, or determiner forms.

Indeed, the conjunction of comprehension and production processes that the PWI paradigm entails may not be ideal for studying gender retrieval during the lexical access of a noun. We have expressed concerns about how this conjunction is the cause of possible mismatches on the time course of target and distractor, which could impose barriers for the observation of effects. Bürki et al. (2020) obtained evidence that confirms these concerns, in that they found that participants' reading times should be measured and included in the models of analysis to properly assess whether there is an influence of the distractor on the production of the target. In this sense, the greatest and most robust effect that has been obtained within the PWI paradigm is that of determiners of different gender. Critically, the flow of activation that reaches determiners comes from different sources, namely, definiteness, number, and gender. Gender has only one source of activation, except in transparent languages, in which a form-based route has been shown to contribute to gender retrieval in most of the nouns. Gender, indeed, seems to be hard to assess even with priming tasks (e.g., Jescheniak, 1999; van Berkum, 1996, 1997), which indicates that the activation of this feature might either be low, have a fast decay rate, or both. Therefore, we should not disregard the possibility that genuine competition between gender nodes may exist but cannot be reflected experimentally, at least not with behavioural measures. This would explain the results of Heim, Eickhoff, et al. (2009) and Heim, Friederici, et al. (2009), who used fMRI measures and found variations on the haemodynamic response related to the factor of gender congruency in areas related to gender processing.

Another way to assess gender selection is during L2 production through a naming task. In this task, the activation of a target noun and its translation emerges through the same source (the concept). Hence, the time course of the flow of activation of both stimuli is very similar, especially in comparison to what occurs with target nouns and written distractor nouns in the PWI paradigm. In the latter, the processing of both stimuli starts at distant points of the lexical access architecture, that is, the target noun is first encoded at the conceptual level and the interfering noun at the lexeme. In naming tasks, both are activated at the conceptual level, increasing the potential duration and intensity of competition. Indeed, when instead of naming tasks, the studies feature translation tasks, some authors have expressed their concerns as to how the time course of lexical activation might be preventing competition between gender nodes to occur, increasing the slipperiness of the cross-linguistic GCE (Bordag & Pechmann,

USC UNIVERSIDADE DE SANTIAGO

2008). In fact, all the null results observed in this area were obtained exclusively with the forward-translation task (Bordag & Pechmann, 2008; Salamoura & Williams, 2007; note that Costa, Kovacic, Franck, et al., 2003 study is controversial). In this task, the interfering noun is processed first at the lexeme level, and it activates its translation (the target noun) at the lemma level. Although closer in time than the stimuli of the PWI paradigm, there is a certain mismatch in comparison to a naming task. Importantly, our meta-analysis did not confirm this rationale based on the time course of naming and translation tasks, as differences in the outcomes were not significant. The tendency, however, would be of higher effects for naming than translation tasks (*g* of .28 vs. .19). A higher number of observations might make it possible to reveal differences between tasks. Hence, this rationale (the time course of naming vs. translation tasks) is not to be dismissed.

In sum, assuming the competitive nature of gender nodes, we believe that the inadequacies in the paradigms, the mismatches on the time course of lexical access of the stimuli involved, and the low degree of activation reaching gender nodes, are directly related to the absence of effects within the PWI paradigm. Yet we also think that there has to be some kind of peculiarity in the bilingual language system that enables the easier observation of gender effects in comparison to paradigms involving native speakers. We think that this peculiarity might lie in the fact that gender nodes in bilinguals normally receive activation from two different lexical entries.

Taking into consideration the bilingual representation of lexical access of BIA-d model (Grainger et al., 2010) and the Multilink model (Dijkstra et al., 2019), the following is expected. Most studies that assessed the cross-linguistic GCE were testing bilinguals who, although unbalanced, had at least moderate proficiency in their L2. This degree of proficiency assures that the amount of activation reaching the gender nodes through the L2 lemmas (or lexemes, if we follow the architecture of the IN model) is not exactly the same but still similar to the one occurring through the L1 lemmas (i.e., links are substantially strong). Not only this, the degree of activation that gender nodes are receiving during language production is rather higher in comparison to monolingual individuals or balanced bilinguals, who do not show so much influence of one language over another. This would suggest that the activation reached by gender nodes is constantly higher in the case of bilinguals (at least those that are not balanced) than in the case of monolinguals. A constant feed of activation is expected to increase the activation baseline of the gender nodes. The effects of competition would hence involve higher

USC UNIVERSIDADE DE SANTIAGO

levels of activation and thus be more consistently obtained. This is exactly what is observed in the bilingual data: small and heterogeneous gender effects, but which are quite consistently observed. The presence of an L2 raises the activation of the gender nodes until levels in which the gender effect is no longer as sensitive to external factors such as the type of task (note the insignificant role of type of task in our meta-analysis). However, if bilinguals are balanced, they should reduce the activation coming from their non-target language (the second source of activation) and the cross-linguistic GCE should eventually disappear. Indeed, this is in line with the results of our third study.

The presence of an L2 within the linguistic system seems to have further repercussions in the observation of effects when considering the role of the language itself. Within the PWI paradigm, the tested language seemed to determine (a) whether effects would be observed when noun phrases were produced (Late Selection Hypothesis), and (b) the type of effects that would be observed with bare nouns (null vs. possibility of lemma competition [GIE]). In the bilingual data, no role was any longer reported for the target language. More precisely, when compared to the literature on native language production, the two main differences regarding the target language in bilingualism are:

(1) Effects of gender and determiner congruency are no longer masked by the gender-form mapping of the elements of agreement, and so they are always obtained with noun phrases even in languages with complex mapping (French and Spanish; Manolescu & Jarema, 2015; Paolieri, Cubelli, et al., 2010; Paolieri et al., 2019).

(2) In bilinguals, the GIE is nowhere to be seen; lemmas connected to the same gender node do not seem to compete for selection.

The reason behind the unmasking of gender effects with noun phrases might be the task itself and the hypothesized higher activation baseline of the gender nodes for unbalanced bilinguals. On the one hand, in the PWI paradigm the interfering noun is read and hence seems to access gender in the first instance. The processes required for the production of the determiner thus have a greater chance of masking the interfering role of the distractor, since the processing of the latter starts somewhat earlier than that of the target. In naming and translation tasks, the closer activation of both stimuli might increase the interfering potential of the "distractor" noun on the target's gender and determiner selection. Hence, the better time-match on lexical access between the target and the interfering noun in bilinguals is one of the possible factors facilitating the observation of the cross-linguistic GCE with noun phrases, even in

USC UNIVERSIDADE DE SANTIAGO

languages with complex gender-form mapping. The other possible explanatory factor is again the higher activation levels reached by gender nodes in bilinguals. These nodes receive activation that adds to an activation baseline that is *per se* higher than that of native speakers. Higher levels of activation might generate greater effects of competition that are more easily obtained in any experimental situation.

Finally, the GIE remains the jigsaw piece that does not seem to fit any puzzle. Indeed, when assuming the non-competitive nature of gender nodes, we found no satisfactory explanation for this effect. Conversely, assuming that gender nodes compete allows us to propose an explanation for the GIE, but we recognise in advance that it is highly tentative. More specifically, if there are indeed bidirectional connections between nouns lemmas/lexemes and gender nodes, it could be that certain individuals can move forward on gender selection and lexical access itself before both lemmas reinforce each other and create a GIE. Other individuals cannot do this. The process of moving forward or surpassing the reinforcement of the lemmas of the same gender would be easier for bilinguals since they are actively working to inhibit the interfering noun, that is, the L1 translation. This would explain the consistent presence of the cross-linguistic GCE, the cross-linguistic GIE which is never seen, as well as the extreme variation of the direction of the potential effects found in our results, acknowledging here the possible existence of individual differences. Indeed, there is evidence in the literature showing that bilinguals differ significantly in their capacity to inhibit their other language during the processing of the target language depending on their inhibitory abilities, as measured by non-linguistic tasks (Linck et al., 2008). Hence, it is sensible to think that the inhibitory capacity of speakers may also create certain individual differences in the processing of a language at different levels of representation.

To summarize, by assuming that gender nodes compete, there is no need for a mixed assumption in which the mechanisms for gender selection vary according to different variables. The causes underlying the differences between the native and the bilingual data on grammatical gender processing would hence be based on the task itself, that is, the insufficiency of the PWI paradigm to detect certain effects of competition, and on the idea that the amount of activation reaching gender nodes is minimal and probably decays rapidly, which hinders the observation of behavioural effects of gender competition.

USC UNIVERSIDADE DE SANTIAGO

7.2 CONTROVERSIAL ISSUES OF THE LITERATURE

Having described the two opposite alternatives that may explain the determiner congruency effect and the gender-based effects observed in the literature and in our own studies, we believe that the validity of both hypotheses directly depends on whether or not future research takes certain aspects into consideration. The first issue to consider is the PWI paradigm itself, which has been revealed as a complex task that relies on the interaction between comprehension and production processes. The literature highlights at least two situations in which this paradigm does not seem to deliver satisfactory results. One of those is that of competition between bound morphemes, for which the simpler singular-plural paradigm consistently finds competitive effects during the selection of bound morphemes. Yet the outcome with the PWI paradigm is mixed and tends not to show them. Another troubling situation is related to the Late Selection Hypothesis. Our results with European Portuguese from Study 2, as well as certain results elsewhere with French participants (Bürki et al., 2020; Foucart et al., 2010), strongly suggest that although effects are null when participants name the pictures using noun phrases in Catalan, Italian, and Spanish, competition for selection is actually occurring. Yet the moment at which the distractor competes with the target occurs too early in time considering the additional processes that have to take place afterwards in these languages for the proper determiner to be selected. Therefore, we are left with absence of evidence in a number of studies that does not seem to constitute evidence for the absence of effects.

When considering such a complex paradigm in which so many processes are involved, certain factors become particularly relevant. Among these, we would highlight stimuli control and sample size. Indeed, it can be especially difficult to assure stimuli control since not only must the target noun be controlled across conditions in different parameters, but also the corresponding picture, the distractor noun, and the relation between the two. Yet it is fundamental to make sure that our outcome is a direct consequence of the process we are studying and not of another process related to other variables. Note that certain studies that lack an adequate degree of control are at the heart of heated discussions, such as Schriefers (1993), which is one of the studies that finds competition between bound morphemes within the PWI paradigm. As seen in Chapter 5, sample sizes are also especially worrying because they directly affect statistical power. Even if underpowered studies must always be avoided, we believe that lack of power is even more of a concern if we are indeed looking at a paradigm that tends to mask effects in a field that also seems to report mostly small effects. For instance, note that all

USC UNIVERSIDADE DE SANTIAGO

studies featuring bare nouns (regardless of the language family) hardly passed the thirty participants mark even though they all used designs consisting of four factors. The possible lack of power puts the reliability of the GIE at stake, as well as the claim that gender is retrieved non-competitively, which is based on only three studies (La Heij et al., 1998; Finocchiaro et al., 2011; Starreveld & La Heij, 2004).

Indeed, the problems we have identified relating to the PWI paradigm itself and the experimental designs that have been used show that opting for one explanatory hypothesis or another can be extremely problematic. Yet, as seen, there is a fundamental problem within the hypothesis assuming the non-competitive nature of grammatical gender, since it has been conceived of and defended on the basis of the absence of effects. Absence of effects should not be interpreted as evidence for an absence per se but simply as absence of evidence. In fact, as noted at the beginning of this section, there are examples of absence of effects in this paradigm that do not entail the non-existence of these effects (e.g., the masking of the "GCE" in Spanish and Italian).

On a similar note, there are certain studies that could have ground-breaking consequences for the way we understand the representation and retrieval of grammatical gender, tipping the balance in favour of one hypothesis or another, yet they require further assessment since they are based on small sample sizes and/or poor stimuli control. Interesting examples are Heim, Friederici, et al. (2009), who used fMRI techniques within the PWI paradigm in German, and Bürki et al. (2016) study, who brought EEG measures to the discussion on gender processing within the PWI paradigm, also in German. Nevertheless, their studies included only 16 and 18 participants, respectively.

Another problem in the literature, and even of this thesis, is the GIE itself, which is systematically null in the multiple attempts that have been carried to replicate it. Regarding this, we would like to highlight two facts. On the one hand, note that our results are in line with those of Finocchiaro et al. (2011), and thus constitute evidence against the GIE. Yet if the GIE is a slippery effect, we cannot disregard the possibility that our results are simply an example of the null results that are expected for heterogeneous effects. On the other hand, an overall view of the studies here makes clear that this a very convoluted area in which every possible situation has been reflected: an effect in opposite directions (the GIE and our unreliable GCE) and null effects. It is thus worth mentioning that this kind of erratic pattern of results (non-replicability of the effects resulting in null effects, plus effects in the opposite direction of what was

USC UNIVERSIDADE DE SANTIAGO

expected; plus effects that have quite a large size when compared to other studies that obtain null results) has been described by many statisticians as a symptom of lack of power (Type S error and overestimation; see Vasishth & Gerlman, 2021). Certainly, lack of power is a cause for the finding of false effects (Brysbaert et al., 2019). In this sense, note that when we updated our meta-analysis on the effects with bare nouns and Romance languages and included our own results, the significant GIE we had observed saw its significance reduced to almost marginal levels (p = .044). Further replication of these experiments will give us the answer we seek regarding the GIE's reliability.

In conclusion, it seems increasingly obvious that we faced with a field that calls desperately for replication, at least if it is our purpose to make well-grounded assumptions on the representation and processing of grammatical gender. The absence of replication has led to there being a lack of studies for the assessment of multiple critical subjects. Yet even if the lack of replication is alarming, and very much needed, we should not make the same mistakes of the past, and certain aspects should be taken into consideration, such as sample sizes, stimuli control, and the use of innovative techniques.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

8 CONCLUDING REMARKS

The aim of the present thesis was to cast light on the convoluted area of grammatical gender representation and processing during language production. To do so, we exhaustively described our object of study – gender –, addressed it from a linguistic and a psycholinguistic point of view, assessed the whole theoretical and experimental literature on the issue, considered together the studies on native language production and bilingualism, and contributed with data of an unexplored language whose peculiarities could bring us closer to the answers we seek. Indeed, our results are more in line with the DSIH within the PWI paradigm, the active role in language production of the routes of gender retrieval proposed by the Dual-Route Model (Gollan & Frost, 2001), as well as the Late Selection Hypothesis. Nevertheless, we obtained robust evidence for genuine gender effects in the forward/backward translation task, but not within the PWI paradigm. We also failed to observe the controversial GIE. Critically, we highlighted the fact that even though we failed to obtain effects of competition between gender values within the PWI paradigm, we cannot interpret these null results as evidence for the idea that gender nodes do not compete for selection, which has been a common interpretation of the absence of gender effects in this field. Rather, that is only one of the possible hypotheses that may explain the whole body of available results and that should be assessed in the future. In fact, when considering the literature from an integrative point of view, we find more satisfactory the explanations sustained by the hypothesis that assumes the competitive nature of gender values but insists on the need of further replication to come to final conclusions.

We believe that, at least partially, we have achieved our goals. Yet we are aware that this thesis has certain limitations that should be taken into account if we want to properly orientate future researchers aiming to make further progress in this area.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

8.1 LIMITATIONS OF THIS THESIS

We would like to start by stating that, although a respectable choice, relying again on the PWI paradigm might seem like we keep stumbling on the same rock. It is undeniable that replication

with new languages, better sample sizes, innovative designs, and better experimental control are all necessary. Still, we should have either accompanied the paradigm with more fine-grained measuring techniques or presented a complementary study with another type of task after our own PWI task replications. It should be noted that experiments with the PWI paradigm testing both European Portuguese speakers as well as speakers of an opaque language, Dutch, while relying on EEGs, are currently on-going in our laboratory. Unfortunately, the COVID-19 pandemic impeded us from finishing and including these experiments for this thesis.

To enhance the outcomes of the PWI paradigm, it might have been useful to include the variable "reading times", as suggested in Bürki et al. (2020), a study which, unfortunately, was published after most of our experiments had been concluded. On a similar note, reading times are specifically prompted by individual differences among participants. We have expressed our concerns regarding the fact that individual differences beyond reading times may impact the pattern of results, but we did not explore this thoroughly. This should be analysed in more detail in future research.

Finally, even though, when compared to previous studies, we used a substantial number of participants and stimuli, our designs were quite complex. They reflected the fact that we had many research questions, ranging from the effects of gender congruency and transparency to the impact of agreement, the gender of the target noun, or animacy. When one has so many research questions, more than 100 participants are usually required to avoid underpowered studies (Brysbaert et al., 2019). Future research should pay special attention to this issue.

8.2 IDEAS FOR FUTURE RESEARCH

Before ending this thesis, we would like to propose a few ideas that might help researchers taking further steps towards arriving at the answers we seek. We believe that the most urgent task in the field is perhaps to conduct further replications of studies with bare nouns, to resolve the debate on the GIE once for all and be sure of the absence of gender effects with Germanic and Slavic languages. Note that since we dealing with a subject in which null results might be a likely expected outcome, it might be wise to conduct preregistered studies. This is mainly because the many null results that may be observed in this field might run the risk of it showing a publication bias, thus further complicating our aim to understanding the processing of grammatical gender during language production. Additionally, these replications should include EEGs measures and allow comparisons between the effects obtained with bare nouns

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

and those obtained with noun phrases. Recording EEGs would allow us to better understand the time course of gender and determiner selection, especially when including extra factors of semantic or phonological relatedness. The inclusion of reading times in the analyses might also be an interesting idea, as it would open new doors in assessing the emergence of gender-based effects with bare nouns. Both EEGs and reading times are also particularly interesting to better evaluate the Late Selection Hypothesis – what constitutes early selection, what processes are actually masking determiner congruency effects in Catalan, Italian, and Spanish, and whether or not this masking can be experimentally surpassed. The replication of studies with bound morphemes, both with the PWI paradigm and the singular-plural paradigm, are also suggested, especially to compare the size of both effects and have new information on the sensitivity of the PWI paradigm to detect potentially smaller effects of competition.

Two of the variables we have given special attention to in this thesis are gender transparency and animacy. Transparency is a variable that – we believe – has remained poorly described in the literature and for which we have made an attempt to develop a more detailed definition. Our ultimate hope is for future authors to conduct classifications of the different languages according to a standardized proposal of transparency measurement such as ours, in order to develop a well-grounded gender transparency continuum (e.g., the one proposed by Kupisch et al., 2018). In this way we would be able to make experimental predictions according to the position of each language on such a continuum. Studies could be conducted in gender acquisition, comprehension, and production according to this classification. In this sense, further attempts to replicate our "GCE" in interaction with ortho-phonological transparency are highly recommended, both with behavioural measures and with EEGs. We need to understand the time course of the routes of gender retrieval in transparent languages and explore whether or not the form-route operates depending on the degree of gender transparency of languages. Regarding animacy, even though our GCE from Experiment 1 in Study 1 was neither robust nor replicated, we believe that the analysis of this variable should not be put aside. There are well-grounded theoretical and empirical concerns which lead us to believe that it might indeed influence the outcomes of any PWI experiment and even the processing of grammatical gender through semantic prioritization. If gender effects are indeed unobtainable through the PWI paradigm, we remind the reader of our suggestion about testing other types of effects (e.g., of phonological overlap, see Study 1) to assess the role of animate nouns within the stimuli list, at least to understand to what extent they affect the attentional mechanisms involved in the task.

USC UNIVERSIDADE DE SANTIAGO

Testing the role of animacy in the determiner congruency effect might be less important, as the requirement for agreement calls for gender processing. Yet if the distractor is indeed being ignored when targets are animate, we might fail to observe the effect.

Finally, bilingual populations offer interesting opportunities to test the tenets of the two alternative hypotheses we presented as a means of explaining the whole set of results. If an unbalanced situation between languages changes the mechanisms of gender selection and suddenly gender selection stops being automatic and turns into a competitive effect – at least temporarily –, or if gender selection is always competitive, we should expect that those bilinguals who show cross-linguistic GCE also show a genuine GCE when considering the PWI paradigm and bare nouns. Note that if gender selection is always competitive, but unbalanced bilinguals receive activation from more lexical sources (L1 and L2) than balanced bilinguals or monolinguals, the genuine GCE should also be expected. This because the lexical entries from both languages are contributing to gender selection and to the proposed rising of the gender activation baseline. It would be interesting to see if this rationale applies and hence if the PWI paradigm can reflect such effects, or on the contrary, whether these effects are not observed and therefore (a) our hypotheses are far from the truth, or (b) not even in this case does the complexity of the PWI paradigm allow effects to be observed, which would be even more alarming. In this sense, we could benefit from studies using EEGs techniques with bilinguals in both a PWI paradigm and a naming task, to understand certain differences between the observed effects (in the case of both noun phrases and bare nouns production; see von Grebmer Zu Wolfsthurn et al., 2021).

8.3 WE ARE ON THE RIGHT PATH

As Stuart Ritchie (2020) claimed when describing a certain unrealistic study that was first praised but later unveiled as a fraud:

> The study fits the exact specifications required to convince Science's peer reviewers it was worth publishing. It was what the publication system and the university job market appeared to demand: not a snapshot of a messy reality where results can be unclear and interpretations uncertain, but a clean, impactful result that could immediately translate to use in the real world. (p. 65)

This thesis presents itself as a snapshot of the messy reality regarding grammatical gender. Indeed, we began our research on grammatical gender amidst mixed results and opposing

theories. We tried to untie a few of the knots that crowd this subject. Had we claimed to untie all the knots that concerned us, our work would have been quite suspicious. Hence, by limiting ourselves to a few reduced goals, we contributed to the field with various kinds of valuable information and ideas that will certainly be helpful for future researchers to work towards untying all those knots. We consider that the type of explorative and unbiased assessment used in this thesis was urgently required in the field. In truth, perhaps the most important take-away of this thesis is that, considering the literature on grammatical gender in its current state, there is no robust conclusion to be established, although the overall data leads us to embrace the idea that the mechanisms underlying gender processing are ones of competition. We further argue that certain previously claimed conclusions might have been grounded in rather controversial inferences or statistically weak results. This view is not presented trivially. We stand by our claim that conducting proper science requires one to take a step back, to observe and analyse the whole body of available work, and to objectively – as possible as one can – judge them. Only by doing so can we truly understand the complexity of a field such as grammatical gender.

It might come as a slight disappointment for the reader to be left with so many open questions. Yet we feel far from disappointed: we are rather excited about the answers that may underlie these questions and about forthcoming research, the future steps of which seem clear to us. After all, we believe that we are on the right path, having established the necessary bases to successfully fulfil this "quest". Certainly, all the questions we have tried to answer throughout this thesis are ultimately linked to the mechanisms underlying gender retrieval in language production. We consider that reaching a more definite conclusion here will be a matter of understanding that these mechanisms are rather flexible, since the nature of grammatical gender representation seems to be driven by language-specific idiosyncrasies, including the number of languages known by the speaker as well as other multiple variables (e.g., transparency and proficiency). Once this broader view on grammatical gender is truly settled in terms of the theoretical position of researchers, they can then explore the field further. This will require the assessment of other languages, the use of more precise and innovative measuring techniques, the improvement of certain methodological aspects, and even perhaps a paradigm shift.

USC UNIVERSIDADE DE SANTIAGO

We do not intend to sound pessimistic here; on the contrary, the field might appear chaotic, but we are grateful that this is the case. Surely, the current state of the art is partly a consequence of many of the bias and replication crises that have recently been unveiled in Psychology (see

Open Science Collaboration, 2015). Yet it is also a reliable reflection of true science. We should never expect research to be a peaceful and immutable process that finds crystalline results and easily answers all the questions it asks. The convoluted mix of results and theories that arise from research into grammatical gender processing during noun production confirms that at least we are now on the right track to find the answers we seek. To prevent this track from deviating we must assure proper replication, avoid publication bias, and better assess slippery and small effects such as those seemingly generated by grammatical gender. We would undoubtedly be more worried if all the philosophers, linguists, and psycholinguists interested on grammatical gender had flattered this feature rather than despised it. Its difficult nature makes its research unavoidably problematic, and research cannot be a bed of roses, as Ritchie's (2020) statement makes clear.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

References

Abutalebi, J., & Green, D. (2007). Bilingual language production: the neurocognition of language representation and control. *Journal of Neurolinguistics*, *20*(30), 242–475. https://doi.org/10.1016/j.jneuroling.2006.10.003

Adelung, J. C. (1783). Von dem Geschlechte der Substantive [About the gender of nouns]. *Magazin für die Deutsche Sprache*, *1*(4), 3–20.

Aikhenvald, A. Y. (2000). *Classifiers: A Typology of Noun Categorization Devices*. Oxford Univesity Press.

Aikhenvald, A. Y., & Mihas, E. I. (2019). *Gender & Classifiers*. Oxford University Press.

Aksenov, A. T. (1984). On the problem of the extralinguistic motivation of grammatical gender. *Voprosy jazykoznanija*, *1*, 14–25.

Alario, F. -X., & Caramazza, A. (2002). The production of determiners: Evidence from French. *Cognition*, *82*(3), 179–223. https://doi.org/10.1016/S0010-0277(01)00158-5

Alario, F. -X., Matos, R. E., & Segui, J. (2004). Gender congruency effects in picture naming. *Acta Psychologica*, *117*(2), 185–204. https://doi.org/10.1016/j.actpsy.2004.06.003

Alderson, J. C. (2006). *Diagnosing foreign language proficiency: The interface between learning and assessment*. Continuum. Retrieved from https://eclass.uoa.gr/modules/document/file.php/ENL286/Testing%20books/Diagnosing%20Foreign%20Language%20Proficiency_Alderson.pdf

Altman, M. N., Khislavsky, A. L., Coverdale, M. E., & Gilger, J. W. (2016). Adaptive attention: how preference for animacy impacts change detection. *Evolution and Human Behavior*, *37*(4), 303–314. https://doi.org/10.1016/j.evolhumbehav.2016.01.006

Altmann, E. M., & Davidson, D. J. (2001). An integrative approach to Stroop: Combining a language model and a unified cognitive theory. *Proceedings of the 23rd Meeting of the Cognitive Science Society* (pp. 21–26). Lawrence Erlbaum Associates, Inc.

Alungum, R. J. C., & Lukas, J. (1978). Some Muhiang grammatical notes. In R. Loving (Ed.), *Miscellaneous papers on Dobu and Arapesh* (pp. 89–130). Summer Institute of Linguistics.

USC DE SANTIAGO DE COMPOSTELA

Alvanoudi, A. (2014). *Grammatical Gender in Interaction: Cultural and cognitive aspects*. Brill.

Arias-Trejo, N., & Alva, E. A. (2013). Early Spanish grammatical gender bootstrapping: Learning nouns through adjectives. *Developmental Psychology, 49*(7), 1308–1314. https://doi.org/10.1037/a0029621

Arias-Trejo, N., Falcón, A., & Alva, E. A. (2013). The gender puzzle: Toddlers' use of articles to access noun information. *Psicológica, 34*(1), 1–23. https://doi.org/10.17575/PSICOLOGIA.V34I1.1435

Aristotle (1984). *The Complete Works of Aristotle. The Revised Oxford Translation*. Princeton University Press.

Aronoff, M. (1994). *Morphology by itself: Stems and inflectional classes*. MIT Press.

Arrivé, M., Gadet, F., & Galmiche, M. (1985). *La grammaire d'aujourd'hui: Guide alphabétique de la langue française* [Grammar today: Alphabetic guide of French]. Paris: Flammarion.

Audring, J. (2008). Gender assignment and gender agreement: Evidence from pronominal gender languages. *Morphology, 18*, 93–116 https://doi.org/10.1007/s11525-009-9124-y

Audring, J. (2009). Reinventing Pronoun Gender [Doctoral dissertation, Vrije Universiteit Amsterdam]. LOT publications. https://www.lotpublications.nl/Documents/227_fulltext.pdf

Audring, J. (2014). Gender as a complex feature. *Language Sciences*, *43*, 5–17. https://doi.org/10.1016/J.LANGSCI.2013.10.003

Audring, J. (2017). Calibrating complexity: How complex is a gender system? *Language Sciences*, *60*, 53–68. https://doi.org/10.1016/j.langsci.2016.09.003

Audring, J. (2019). Canonical, complex, complicated? In F. Di Garbo, B. Olsson & B. Wälchli (Eds.), *Grammatical gender and linguistic complexity: Volume I: General issues and specific studies (Studies in Diversity Linguistics 26)* (pp. 15–52). Language Science Press.

Ayora, P., Peressotti, F., Alario, F.-X., Mulatti, C., Pluchino, P., Job, R., & Dell'Acqua, R. (2011). What Phonological Facilitation Tells about Semantic Interference: A Dual-Task Study. *Frontiers in Psychology*, *2*. https://doi.org/10.3389/fpsyg.2011.00057

USC UNIVERSIDADE DE SANTIAGO

Ayoun, D. (2018). Grammatical gender assignment in French: Dispelling the native speaker myth. *Journal of French Language Studies*, *28*(1), 113–148. https://doi.org/10.1017/S095926951700014X

Azeredo, J. C. d. (2013). *Gramática Houaiss da Língua Portuguesa* [Houaiss Grammar of the Portuguese Language]. Publifolha.

Baayen, R. H. (2008). Analyzing linguistic data: A practical introduction to statistics using R. Cambridge University Press.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Badecker, W., Miozzo, M., & Zanuttini, R. (1995). The two-stage model of lexical retrieval: evidence from a case of anomia with selective preservation of grammatical gender. *Cognition*, *57*(2), 193–216. https://doi.org/10.1016/0010-0277(95)00663-j

Bally, C. (1935). *Le langage et la vie* [The language and the life]. *2nd ed*. Max Niehans Éditeur.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates D., Maechler M., Bolker B., & Walker S. (2014). *lme4: Linear Mixed-Effects Models Using Eigen and S4. R package version*, *1*, 1–10. Available online at: http://CRAN.R-project.org/package=lme4

Bates, E., Devescovi, A., Hernandez, A., & Pizzamiglio, L. (1996). Gender priming in Italian. *Perception & Psychophysics*, *58*(7), 992–1004. https://doi.org/10.3758/bf03206827

Bates, E., Devescovi, A., Pizzamiglio, L., D'Amico, S., & Hernandez, A. (1995). Gender and lexical access in Italian. *Perception & Psychophysics*, *57*(6), 847–862. https://doi.org/10.3758/bf03206800

Bates, D., Maechler, M, Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Beckett, M. (2010). *Gender Assignment and Word-final Pronunciation in French: Two semantic systems*. Lincom.

USC UNIVERSIDADE DE SANTIAGO

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

Bermúdez-Otero, R. (2008). *A tutorial on lexical classes*. From Workshop on Theoretical Morphology 4. Retrieved from http://www.bermudez-otero.com/WOTM4.pdf

Bi, Y., Wei, T., Janssen, N., & Han, Z. (2009). The contribution of orthography to spoken word production: evidence from Mandarin Chinese. *Psychonomic Bulletin & Review*, *16*(3), 555–560. https://doi.org/10.3758/PBR.16.3.555

Bidot, E. (1925). *La clef du genre des substantifs français (méthode dispensant d'avoir recours au dictionnaire)* [The key to the gender of French nouns (method dispensing from having to resort to the dictionary)]. Imprimerie Nouvelle.

Bloomfield, L. (1933). *Language*. Holt, Rinehart & Winston.

Boersma, P., & Weenink, D. (2018). *Praat: doing phonetics by computer* [Computer program]. Retrieved from http://www.praat.org

Boloh, Y., Escudier, F., Royer, S., & Ibernon, L. (2012) Determiners and noun endings in French children's gender attribution choices. *First Language*, *32*(3), 398–410. https://doi.org/10.1177/0142723711422621

Boloh, Y., & Ibernon, L. (2010). Gender attribution and gender agreement in 4- to 10-year-old French children. *Cognitive Development*, *25*(1), 1–25. https://doi.org/10.1016/j.cogdev.2009.09.011

Boloh, Y., & Ibernon, L. (2013). Natural gender, phonological cues and the default grammatical gender in French children. *First Language*, *33*(5), 449–468. https://doi.org/10.1177/0142723713499848

Bölte, J., & Connine, C. M. (2004). Grammatical gender in spoken word recognition in German. *Perception & Psychophysics*, *66*(6), 1018–1032. https://doi.org/10.3758/bf03194992

Bonfieni, M., Branigan, H. P., Pickering, M. J., & Sorace, A. (2019). Language experience modulates bilingual language control: The effect of proficiency, age of acquisition, and exposure on language switching. *Acta Psychologica*, *193*, 160–170. https://doi.org/10.1016/j.actpsy.2018.11.004

Bonin, P., Gelin, M., & Bugaiska, A. (2014). Animates are better remembered than inanimates: further evidence from word and picture stimuli. *Memory & Cognition*, *42*(3), 370–382. https://doi.org/10.3758/s13421-013-0368-8

Bonin, P., Gelin, M., Dioux, V., & Méot, A. (2019). “It is alive!” Evidence for animacy effects in semantic categorization and lexical decision. *Applied Psycholinguistics*, *40*(4), 965–985. https://doi.org/10.1017/s0142716419000092

Bordag, D. (2004). Interaction of L1 and L2 systems at the level of grammatical encoding: Evidence from picture naming. *EUROSLA Yearbook, 4*, 203–230. https://doi.org/10.1075/eurosla.4.10int

Bordag, D., & Pechmann, T. (2007). Factors influencing L2 gender processing. *Bilingualism: Language and Cognition*, *10*, 299–314. https://doi.org/10.1017/s1366728907003082

Bordag, D., & Pechmann, T. (2008a). Grammatical gender in speech production: Evidence from Czech. *Journal of Psycholinguistic Research*, *37*(2), 69–85. https://doi.org/10.1007/s10936-007-9060-0

Bordag, D., & Pechmann, T. (2008b). Grammatical gender in translation. *Second Language Research, 24*, 139–166. https://doi.org/10.1177/0267658307086299

Borenstein, M. (2009). Effect sizes for continuous data. In H. Cooper, L. Hedges & J. C. Valentine (Eds.), *The Handbook of Research Synthesis and Meta-analysis (2nd Ed.)* (pp. 221–235). Russell Sage Foundation.

Bosquart, M. (1998). *Nouvelle grammaire* [New grammar]. Guérin.

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

Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations (with discussion). *Journal of the Royal Statistical Society: Series B (Methodological)*, *26*(2), 211–252. https://doi.org/10.1111/j.2517-6161.1964.tb00553.x

Branigan, H. P., Pickering, M. J., & Tanaka, M. (2008). Contributions of animacy to grammatical function assignment and word order during production. *Lingua*, *118*(2), 172–189. https://doi.org/10.1016/j.lingua.2007.02.003

Brugmann, K. (1889). Das Nominalgeschlecht in den indogermanischen Sprachen [The nominal gender in the Indo-European languages]. Internationale Zeitschrift für allgemeine Sprachwissenschaft 4, 100–109.

Brysbaert, M., Buchmeier, M., Conrad, M., Jacobs, A. M., Bölte, J., & Böhl, A. (2011). The word frequency effect: A review of recent developments and implications for the choice

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

of frequency estimates in German. *Experimental Psychology, 58*(5), 412–424. https://doi.org/10.1027/1618-3169/a000123

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods, 41,* 977–990. https://doi.org/10.3758/BRM.41.4.977

Brysbaert, M. (2019). How Many Participants Do We Have to Include in Properly Powered Experiments? A Tutorial of Power Analysis with Reference Tables. *Journal of Cognition, 2*(1): 16, 1–38. https://doi.org/10.5334/joc.72

Bultena, S., Dijkstra, T., & Van Hell, J. (2015). Language switch costs in sentence comprehension depend on language dominance: Evidence from self-paced reading. *Bilingualism: Language and Cognition*, *18*(3), 453–469. https://doi.org/10.1017/S1366728914000145

Burke, L. A., & James, K. E. (2006). Using online surveys for primary research data collection: lessons from the field. *Journal of Innovation and Learning*, *3*(1), 16–30. https://doi.org/10.1504/IJIL.2006.008177

Bürki, A., Besana, T., Degiorgi, G., Gilbert, R., & Alario, F. -X. (2020). Representation and selection of determiners with phonological variants. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(7), 1287–1315. https://doi.org/10.1037/xlm0000643

Bürki, A., Sadat, J., Dubarry, A.-S., & Alario, F. -X. (2016). Sequential processing during noun phrase production. *Cognition*, *146*, 90–99. https://doi.org/10.1016/j.cognition.2015.09.002

Caffarra, S., & Barber, H. A. (2015). Does the ending matter? The role of gender-to-ending consistency in sentence reading. *Brain Research*, *1605*, 83–92. https://doi.org/10.1016/j.brainres.2015.02.018

Caffarra, S., Barber, H., Molinaro, N., & Carreiras, M. (2017). When the end matters: influence of gender cues during agreement computation in bilinguals. *Language, Cognition and Neuroscience*, *32*(9), 1069–1085. https://doi.org/10.1080/23273798.2017.1283426

Caffarra, S., Janssen, N., & Barber, H. A. (2014). Two sides of gender: ERP evidence for the presence of two routes during gender agreement processing. *Neuropsychologia*, *63*, 124–134. https://doi.org/10.1016/j.neuropsychologia.2014.08.016

USC UNIVERSIDADE DE SANTIAGO

Caffarra, S., Siyanova-Chanturia, A., Pesciarelli, F., Vespignani, F., & Cacciari, C. (2015). Is the noun ending a cue to grammatical gender processing? An ERP study on sentences in Italian. *Psychophysiology*, *52*(8), 1019–1030. https://doi.org/10.1111/psyp.12429

Câmara Jr., J. M. (1970). *Estrutura da Língua Portuguesa*. Vozes.

Caramazza, A. (1997). How Many Levels of Processing Are There in Lexical Access? *Cognitive Neuropsychology*, *14*(1), 177–208. https://doi.org/10.1080/026432997381664

Caramazza, A., & Costa, A. (2000). The semantice interference effect in the picture–word interference paradigm: Does the response set matter? *Cognition*, *75*(2), B51–B64. https://doi.org/10.1016/S0010-0277(99)00082-7

Caramazza, A., & Miozzo, M. (1997). The relation between syntactic and phonological knowledge in lexical access: Evidence from the "tip-of-the-tongue" phenomenon. *Cognition*, *64*(3), 309–343. https://doi.org/10.1016/S0010-0277(97)00031-0

Caramazza, A., & Miozzo, M. (1998). More is not always better: a response to Roelofs, Meyer, and Levelt. *Cognition*, *69*(2), 231–241. https://doi.org/10.1016/s0010-0277(98)00057-2

Caramazza, A., Miozzo, M., Costa, A., Schiller, N. O., & Alario, F.-X. (2001). A Cross-linguistic Investigation of Determiner Production. In E. Dupoux (Ed.), *Language, brain, and cognitive development: Essays in honor of Jacques Mehler* (pp. 209–226). MIT Press.

Carello, C., Lukatela, G., & Turvey, M. T. (1988). Rapid naming is affected by association but not by syntax. *Memory & Cognition*, *16*(3), 187–195. https://doi.org/10.3758/BF03197751

Carroll, S. (1989). Second-language acquisition and the computational paradigm. *Language Learning*, *39*, 535–594. https://doi.org/10.1111/j.1467-1770.1989.tb00902.x

Caselli, M. C., Leonard, L. B., Volterra, V., & Campagnoli, M.G. (1993). Toward mastery of Italian morphology: a cross-sectional study. *Journal of Child Language*, *20*(2), 377–393. https://doi.org/10.1017/S0305000900008333

Chomsky, N. (1995) *The Minimalist Program*. The MIT Press.

Colé, P., Pynte, J., & Andriamamonjy, P. (2003). Effect of grammatical gender on visual word recognition: Evidence from lexical decision and eye movement experiments. *Perception & Psychophysics*, *65*, 407–419. https://doi.org/10.3758/BF03194572

USC UNIVERSIDADE DE SANTIAGO

Comesaña, M., Bertin, P., Oliveira, H., Soares, A. P., Hernández-Cabrera, J. A., & Casalis, S. (2018). The impact of cognateness of word bases and suffixes on morpho-orthographic processing: A masked priming study with intermediate and high-proficiency Portuguese-English bilinguals. *PLOS One*, *13*(3), e0193480. https://doi.org/10.1371/journal.pone.0193480

Comesaña, M., Ferré, P., Romero, J., Guasch, M., Soares, A. P., & García-Chico, T. (2015). Facilitative effect of cognate words vanishes when reducing the orthographic overlap: The role of stimuli list composition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(3), 614–635. https://doi.org/10.1037/xlm0000065

Corbett, G. G. (1978). The agreement hierarchy. *Journal of Linguistics*, *15*(2), 203–224. https://doi.org/10.1017/S0022226700016352

Corbett, G. G. (1991). *Gender*. Cambridge University Press. https://doi.org/10.1017/CBO9781139166119

Corbett, G. G. (1994). Gender and gender systems. In R. E. Asher & J. M. Y. Simpson (Eds.), *The Encyclopedia of Language and Linguistics I* (pp. 1347–1353). Pergamon Press.

Corbett, G. G. (2000). *Number*. Cambridge University Press. https://doi.org/10.1017/CBO9781139164344

Corbett, G. G. (2006) *Agreement*. Cambridge University Press.

Corbett, G. G. (2009). Gender and noun classes. In T. Shopen (Ed.), *Language Typology and Syntactic Description* (pp. 241–279). Cambridge University Press. https://doi.org/10.1017/CBO9780511618437.004

Corbett, G. G. (2012). *Features*. Cambridge University Press. https://doi.org/10.1017/CBO9781139206983

Corbett. G. G. (2013a). Number of Genders. In M. S. Dryer & M. Haspelmath (Eds.), *The World Atlas of Language Structures Online*. Max Planck Institute for Evolutionary Anthropology. Retrieved from http://wals.info/chapter/30

Corbett. G. G. (2013b). Sex-based and Non-sex-based Gender Systems. In M. S. Dryer & M. Haspelmath (Eds.), *The World Atlas of Language Structures Online.* Max Planck Institute for Evolutionary Anthropology. Retrieved from http://wals.info/chapter/31

Corbett, G. G., & Fedden, S. (2015). Canonical gender. Journal of Linguistics, 52(3), 495–531. https://doi.org/10.1017/S0022226715000195

USC DE SANTIAGO DE COMPOSTELA

Corrêa, L. M. S., Augusto, M. R. A., & Castro, A. (2011). Agreement and markedness in the ascription of gender to novel animate nouns by children acquiring Portuguese. *Journal of Portuguese Linguistics*, *10*(1), 121–142. https://doi.org/10.5334/jpl.103

Corrêa, L. M. S., & Name, C. (2003). The Processing of Determiner-Noun Agreement and the Identification of the Gender of Nouns in the Early Acquisition of Portuguese. *Journal of Portuguese Linguistics*, *2*, 19–43. https://doi.org/10.5334/jpl.34

Corteen, E. (2019). *The Assignment of Grammatical Gender in German: Testing Optimal Gender Assignment Theory* [Doctoral dissertation, University of Cambridge]. Apollo. https://doi.org/10.17863/CAM.37638

Costa, A., Kovacic, D., Fedorenko, E., & Caramazza, A. (2003). The Gender Congruency Effect and the Selection of Freestanding and Bound Morphemes: Evidence From Croatian. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(6), 1270–1282. https://doi.org/10.1037/0278-7393.29.6.1270

Costa, A., Kovacic, D., Franck, J., & Caramazza, A. (2003). On the autonomy of the grammatical gender systems of the two languages of a bilingual. *Bilingualism: Language and Cognition*, *6*(3), 181–200. https://doi.org/10.1017/s1366728903001123

Costa, A., Sebastián-Gallés, N., Miozzo, M., & Caramazza, A. (1999). The gender congruity effect: Evidence from Spanish and Catalan. *Language and Cognitive Processes*, *14*(4), 381–391. https://doi.org/10.1080/016909699386275

Costa, A. S., Comesaña, M., & Soares, A.P. (2021). *PHOR-in-One: A Multilingual Lexical Database With PHonological, ORthographic and Phonographic Word Similarity Estimates in Four Languages* [Manuscript in preparation]. School of Psychology, University of Minho.

Cubelli, R., Lotto, L., Paolieri, D., Girelli, M., & Job, R. (2005). Grammatical gender is selected in bare noun production: Evidence from the picture–word interference paradigm. *Journal of Memory and Language*, *53*(1), 42–59. https://doi.org/10.1016/j.jml.2005.02.007

Curzan, A. (2003). *Gender shifts in the history of English*. Cambridge University Press. https://doi.org/10.1017/CBO9780511486913

Dahl, Ö. (2000). Animacy and the notion of semantic gender. In B. Unterbeck, M. Rissanen, T. Nevalainen, & M. Saari (Eds.), *Gender in Grammar and Cognition*, (pp. 99–116). De Gruyter Mouton. https://doi.org/10.1515/9783110802603.99

USC UNIVERSIDADE DE SANTIAGO

Dank, M., & Deutsch, A. (2015). Morphological structure governs the process of accessing grammatical gender in the course of production. *The Mental Lexicon*, *10*(2), 186–220. https://doi.org/10.1075/ml.10.2.02dan

Dank, M., Deutsch, A., & Bock, K. (2015). Resolving Conflicts in Natural and Grammatical Gender Agreement: Evidence from Eye Movements. *Journal of Psycholinguistic Research*, *44*(4), 435–467. https://doi.org/10.1007/s10936-014-9291-9

Davies, R., Cuetos, F., & Glez-Seijas, R. M. (2007). Reading development and dyslexia in a transparent orthography: a survey of Spanish children. *Annals of Dyslexia*, *57*(2), 179–198. https://doi.org/10.1007/s11881-007-0010-1

de Courtenay, B. (1929). Einfluß der Sprache auf Weltanschauung und Stimmung [Influence of language on worldview and mood]. *Prace Filologiczne*, *14*, 185–256.

De Martino, M., Bracco, G., & Laudanna, A. (2011). The activation of grammatical gender information in processing Italian nouns. *Language and Cognitive Processes*, *26*(4-6), 745–776. https://doi.org/10.1080/01690965.2010.491977

De Martino, M., Bracco, G., Postiglione, F., & Laudanna, A. (2017). The influence of grammatical gender and suffix transparency in processing Italian written nouns. *The Mental Lexicon*, *12*(1), 107–128. https://doi.org/10.1075/ml.12.1.05dem

Defior, S., Martos, F., & Cary, L. (2002). Differences in reading acquisition development in two shallow orthographies: Portuguese and Spanish. *Applied Psycholinguistics*, *23*(1), 135–148. https://doi.org/10.1017/S0142716402000073

de Leeuw J. R. (2015). jsPsych: a JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods*, *47*(1), 1–12. https://doi.org/10.3758/s13428-014-0458-y

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, *93*(3), 283–321. https://doi.org/10.1037/0033-295x.93.3.283

Dell, G. S. (1990). Effects of Frequency and Vocabulary Type on Phonological Speech Errors. *Language and Cognitive Processes*, *5*(4), 313–349. https://doi.org/10.1080/01690969008407066

Dell, G. S., & Reich, P. A. (1981). Stages in sentence production: An analysis of speech error data. *Journal of Verbal Learning & Verbal Behavior*, *20*(6), 611–629. https://doi.org/10.1016/S0022-5371(81)90202-4

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Desrochers, A., & Brabant, M. (1995). Interaction entre facteurs phonologiques et sémantiques dans une épreuve de catégorisation lexicale [Interaction between phonological and semantic factors in a lexical categorization test]. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, *49*(2), 240–262. https://doi.org/10.1037/1196-1961.49.2.240

Desrochers, A., & Paivio, A. (1990). Le phonème initial des noms inanimés et son effet sur l'identification du genre grammatical [The initial phoneme of inanimate nouns and its effect on grammatical gender identification.]. *Canadian Journal of Psychology*, *44*, 44–57.

Desrochers, A., Paivio, A., & Desrochers, S. (1989). L'effet de la fréquence d'usage des noms inanimés et de la valeur prédictive de leur terminaison sur l'identification du genre grammatical [The effect of frequency of use of inanimate nouns and the predictive value of their termination on the identification of grammatical gender]. *Revue Canadienne de Psychologie*, *43*, 62–73.

Deutsch, A., & Dank, M. (2018). Morphological structure mediates the notional meaning of gender marking: Evidence from the gender-congruency effect in Hebrew speech production. *Quarterly Journal of Experimental Psychology*, *72*(3), 389–402. https://doi.org/10.1177/1747021818757942

Deutsch, W., & Wijnen, F. (1985). The article's noun and the noun's article: explorations into the representation and access of linguistic gender in Dutch. *Linguistics*, *23*, 793–810. https://doi.org/10.1515/ling.1985.23.5.793

Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology*, *5*, 781. https://doi.org/10.3389/fpsyg.2014.00781

Dijkstra, T., Miwa, K., Brummelhuis, B., Sappelli, M., & Baayen, H. (2010). How crosslanguage similarity and task demands affect cognate recognition. *Journal of Memory and Language*, *62*, 284–301. https://doi.org/10.1016/j.jml.2009.12.003

Dijkstra, T., Wahl, A., Buytenhuijs, F., Van Halem, N., Al-Jibouri, Z., De Korte, M., & Rekké, S. (2019). Multilink: A computational model for bilingual word recognition and word translation. *Bilingualism: Language and Cognition*, *22*(4), 657–679. https://doi.org/10.1017/S1366728918000287

USC UNIVERSIDADE DE SANTIAGO

Dimitropoulou, M., Duñabeitia, J. A., Avilés, A., Corral, J., & Carreiras, M. (2010). Subtitle-based word frequencies as the best estimate of reading behavior: the case of greek. *Frontiers in psychology*, *1*, 218. https://doi.org/10.3389/fpsyg.2010.00218

Dixon, R. M. W. (1972). *The Dyirbal Language of North Queensland*. Cambridge University Press. https://doi.org/10.1017/CBO9781139084987

Doleschal, U. (2014). Nominale Kategorien: Genus. In K. Gutschmidt, S. Kempgen, T. Berger & P. Kosta (Eds.), *Die slavischen Sprachen: Ein internationales Handbuch zu ihrer Struktur, ihrer Geschichte und ihrer Erforschung* [The Slavic languages: An international handbook on their structure, their history and their investigation] (pp. 143–152). De Gruyter Mouton.

Dumay, N., Frauenfelder, U. H., & Content, A. (2002). The role of the syllable in lexical segmentation in French: word-spotting data. *Brain and language*, *81*(1-3), 144–161. https://doi.org/10.1006/brln.2001.2513

Dunlap, W. P., Cortina, J. M., Vaslow, J. B., & Burke, M. J. (1996). Meta-analysis of experiments with matched groups or repeated measures designs. *Psychological Methods*, *1*(2), 170–177. https://doi.org/10.1037/1082-989X.1.2.170

Dussias, P. E., & Sagarra, N. (2007). The effect of exposure on syntactic parsing in Spanish-English bilinguals. *Bilingualism: Language and Cognition*, *10*(1), 101–116. https://doi.org/10.1017/S1366728906002847

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. https://doi.org/10.1177/1367006911435719

Faiß, K. (1989). *Englische Sprachgeschichte* [English language history]. Tübingen.

Fedden, S., & Corbett, G. G. (2017) Gender and classifiers in concurrent systems: Refining the typology of nominal classification. *Glossa: A journal of general linguistics*, *2*(1), 1–47. https://doi.org/10.5334/gjgl.177

Félix, S. B., Pandeirada, J. N. S., & Nairne, J. S. (2019). Adaptive memory: longevity and learning intentionality of the animacy effect. *Journal of Cognitive Psychology*, *31*(3), 251–260. https://doi.org/10.1080/20445911.2019.1586716

Fernández-Fuertes, R., Álvarez-de-la-Fuente, E., & Mujcinovic, S. (2016). The acquisition of grammatical gender in L1 bilingual Spanish. In A. A. de-la-Fuente, E. Valenzuela & C.

Martínez-Sanz (Eds.), *Language Acquisition Beyond Parameters: Studies in honour of Juana M. Liceras* (pp. 237–279). https://doi.org/10.1075/sibil.51.10fue

Figueira, R. A. (2001). Marcas insólitas na aquisição do gênero gramatical: a propriedade reflexiva da linguagem na fala da criança. *Letras de Hoje*, *36*(3), 313–320. Retrieved from https://revistaseletronicas.pucrs.br/ojs/index.php/fale/article/view/14582

Finocchiaro, C., Alario, F. X., Schiller, N. O., Costa, A., Miozzo, M., & Caramazza, A. (2011). Gender congruency goes Europe: A cross-linguistic study of the gender congruency effect in Romance and Germanic languages. *Rivista di Linguistica*, *23*(2), 161–198.

Finocchiaro, C., & Caramazza, A. (2006). The production of pronominal clitics: Implications for theories of lexical access. *Language and Cognitive Processes*, *21*(1-3), 141–180. https://doi.org/10.1080/01690960400001887

Fisher, Z., & Tipton, E. (n.a.). *URL: robumeta: An R-package for robust variance estimation in meta-analysis*. URL: https://cran.r-project.org/web/packages/robumeta/vignettes/robumetaVignette.pdf

Fisher, Z., Tipton, E., & Zhipeng, H. (2017). *robumeta: Robust Variance Meta-Regression. R package version 2.0*. URL: https://CRAN.R-project.org/package=robumeta

Fodor, I. (1959). The origin of grammatical gender. *Lingua*, *8*, 186–214. https://doi.org/10.1016/0024-3841(59)90020-8

Foris, D. P. (2000). *A grammar of Sochiapam Chinantec: Studies in Chinantec Languages 6*. SIL International and the University of Texas at Arlington.

Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods*, *35*(1), 116–124. https://doi.org/10.3758/bf03195503

Foucart, A., Branigan, H. P., & Bard, E. G. (2010). Determiner selection in romance languages: Evidence from French. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(6), 1414–1421. https://doi.org/10.1037/a0020432

Frota, S., Butler, J., Correia, S., Severino, C., Vicente, S., & Vigário, M. (2015). *Questionários McArthur-Bates CDI para o Português Europeu: formas reduzidas*. Laboratório de Fonética, CLUL/FLUL. Retrieved from http://labfon.letras.ulisboa.pt/babylab/pt/CDI

Frota, S., Correia, S., Severino, C., Cruz, M., Vigário, M., & Cortês, S. (2012). *PLEX5 - A production lexicon of child speech for European Portuguese*. Laboratório de Fonética CLUL/FLUL.

USC UNIVERSIDADE DE SANTIAGO

Frota, S., Vigário, M., Martins, F., & Cruz, M. (2010). *FrePOP. Version 1.0. Laboratório de Fonética (CLUL)*. Faculdade de Letras da Universidade de Lisboa. Retrieved from http://frepop.fl.ul.pt/

Fung, K. T. D., & Murphy, V. A. (2016). Cross linguistic influence in adult L2/L3 learners: The case of French on English morphosyntax. *Journal of Education*, *3*, 6–15. https://doi.org/10.7603/s40742-015-0011-4

Garrett, M. F. (1980). Levels of processing in sentence production. In B. Butterworth (Ed.), *Language Production: Volume 1*. Academic Press.

Garrett, M. F. (1992). Disorders of lexical selection. *Cognition*, *42*(1-3), 143–180. https://doi.org/10.1016/0010-0277(92)90042-G

Ghosh, A. (2008). Santali. In G. S. Anderson (Ed.), *The Munda languages* (pp. 11–98). Routledge. https://doi.org/10.4324/9781315822433.chThree

Girolami-Boulinier, A., & Pinto, M. d. G. (1994). A ortografia em crianças francesas, inglesas e portuguesas [The orthography in French, English, and Portuguese children]. *Revista da Faculdade de Letras: Línguas e Literaturas*, *11*, 99–113. Retrieved from http://aleph.letras.up.pt/index.php/rll/article/view/8322/7612

Gollan, T. H., & Frost, R. (2001). Two routes to grammatical gender: Evidence from Hebrew. *Journal of Psycholinguistic Research*, *30*(6), 627–651. https://doi.org/10.1023/A:1014235223566

Goswami, U., Gombert, J. E., & de Barrera, L. F. (1998). Children's orthographic representations and linguistic transparency: Nonsense word reading in English, French, and Spanish. *Applied Psycholinguistics*, *19*(1), 19–52. https://doi.org/10.1017/S0142716400010560

Grainger, J., & Dijkstra, T. (1992). On the representation and use of language information in bilinguals. *Advances in Psychology*, *83*, 207–220. https://doi.org/10.1016/s0166-4115(08)61496-x

Grainger, J., Midgley, K., & Holcomb, P. J. (2010). Chapter 14. Re-thinking the bilingual interactive-activation model from a developmental perspective (BIA-d). In M. Kail & M. Hickmann, *Language Acquisition Across Linguistic and Cognitive Systems* (pp. 267–283). https://doi.org/10.1075/lald.52.18gra

USC UNIVERSIDADE DE SANTIAGO

Grainger, J., O'Regan, J. K., Jacobs, A. M., & Segui, J. (1989). On the role of competing word units in visual word recognition: the neighbourhood frequency effect. *Perception & Psychophysics*, *45*(3), 189–195. https://doi.org/10.3758/bf03210696

Greenberg, J. H. (1977). Numeral classifiers and substantival number: problems in the genesis of a linguistic type. In A. Makkai, V. Becker & L. Heilmann (Eds.), *Linguistics at the crossroads* (pp. 276–300). Bologna Liviana Editrice.

Grinevald, C. (2000). A morphosyntactic typology of classifiers. In J. Senft (Ed.), *Nominal Classification* (pp. 50–92). Cambridge University Press.

Guasch, M., Boada, R., Ferré, P., & Sánchez-Casas, R. (2013). NIM: A Web-based Swiss Army knife to select stimuli for psycholinguistic studies. *Behavior Research Methods*, *45*(3), 765–771. https://doi.org/10.3758/s13428-012-0296-8

Guillelmon, D., & Grosjean, F. (2001). The gender marking effect in spoken word recognition: the case of bilinguals. *Memory & Cognition*, *29*(3), 503–511. https://doi.org/10.3758/bf03196401

Guo, T., Liu, H., Misra, M., & Kroll, J. F. (2011). Local and global inhibition in bilingual word production: fMRI evidence from Chinese-English bilinguals. *NeuroImage*, *56*(4), 2300–2309. https://doi.org/10.1016/j.neuroimage.2011.03.049

Hardison, D. (1996). Gender assignment to nonwords in French: implications for the role of the final syllable in lexical processing and organization of the mental lexicon. In *Proceedings of the Indiana Linguistics Club Twenty-Fifth Anniversary Volume* (pp. 13–25). Indiana University Linguistics Club.

Harris, J. W. (1980). Nonconcatenative Morphology and Spanish Plurals. *Journal of Linguistic Research, 1*, 14–31.

Harris, J. W. (1991). The Exponence of Gender in Spanish. *Linguistic Inquiry*, *22*(1), 27–62. https://doi.org/10.1007/978-94-011-2516-1_6

Hartsuiker, R. J., & Kolk, H. H. J. (2001). Error monitoring in speech production: A computational test of the perceptual loop theory. Cognitive Psychology, *42*(2), 113–157. https://doi.org/10.1006/cogp.2000.0744

Hawkins, R., & Franceschina, F. (2004). Explaining the Acquisition and Non-acquisition of Determiner-noun Gender Concord in French and Spanish. In J. Paradis & P. Preüvost (Eds.), *The Acquisition of French in Different Contexts* (pp. 175–205). John Benjamins.

USC UNIVERSIDADE DE SANTIAGO

Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics*, *6*(2), 107–128. https://doi.org/10.2307/1164588

Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Erratum: Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods*, *1*(2), 164–165. https://doi.org/10.1002/jrsm.17

Heidolph, K. E., Flämig, W., & Motsch, W. (1984). *Grundzüge einer deutschen Grammatik* [Basics of a German grammar]. Akademie-Verlag, 2. Auflage.

Heim, S., Alter, K., & Friederici, A. D. (2005). A dual-route account for access to grammatical gender: evidence from functional MRI. *Anatomy and embryology*, *210*(5-6), 473–483. https://doi.org/10.1007/s00429-005-0032-6

Heim, S., Eickhoff, S. B., Friederici, A. D., & Amunts, K. (2009). Left cytoarchitectonic area 44 supports selection in the mental lexicon during language production. *Brain Structure and Function*, *213*(4-5), 441–456. https://doi.org/10.1007/s00429-009-0213-9

Heim, S., Friederici, A. D., Schiller, N. O., Rüschemeyer, S. A., & Amunts, K. (2009). The determiner congruency effect in language production investigated with functional MRI. *Human brain mapping*, *30*(3), 928–940. https://doi.org/10.1002/hbm.20556

Heim, S., Opitz, B., & Friederici, A. D. (2002). Broca's area in the human brain is involved in the selection of grammatical gender for language production: evidence from event-related functional magnetic resonance imaging. *Neuroscience letters*, *328*(2), 101–104. https://doi.org/10.1016/s0304-3940(02)00494-9

Heim, S., Opitz, B., Müller, K., & Friederici, A. D. (2003). Phonological processing during language production: fMRI evidence for a shared production-comprehension network. Brain research. *Cognitive brain research*, *16*(2), 285–296. https://doi.org/10.1016/s0926-6410(02)00284-7

Hellinger, M., & Bußmann, H. (2001). *Gender across languages: The linguistic representation of women and men, Vol. 1*. John Benjamins Publishing Company. https://doi.org/10.1075/impact.9

Henaff Gonon, M. A., Bruckert, R., & Michel, F. (1989). Lexicalization in an anomic patient. *Neuropsychologia*, *27*(4), 391–407. https://doi.org/10.1016/0028-3932(89)90047-x

USC UNIVERSIDADE DE SANTIAGO

Hernandez, A. E., Kotz, S. A., Hofmann, J., Valentin, V. V., Dapretto, M., & Bookheimer, S. Y. (2004). The neural correlates of grammatical gender decisions in Spanish. *Neuroreport, 15*(5), 863–866. https://doi.org/10.1097/00001756-200404090-00026

Hickey, R. (1999). The phonology of gender in Modern German. In M. Rissanen and B. Unterbeck (Eds), *Gender. Cross-linguistic Studies* (pp. 621–663). De Gruyter Mouton.

Hockett, C. F. (1958). *A course in modern linguistics*. Holt/Rinehart and Winston. https://doi.org/10.1111/j.1467-1770.1958.tb00870.x

Hohlfeld, A. (2006). Accessing grammatical gender in German: The impact of gender-marking regularities. *Applied Psycholinguistics*, *27*(2), 127–142. https://doi.org/10.1017/S0142716406060218

Holmes, V. M., & de la Bâtie, B. D. (1999). Assignment of grammatical gender by native speakers and foreign learners of French. *Applied Psycholinguistics, 20*(4), 479–506. https://doi.org/10.1017/S0142716499004026

Holmes, V. M., & Segui, J. (2004). Sublexical and lexical inXuences on gender assignment in French. *Journal of Psycholinguistic Research, 33*, 425–457. https://doi.org/10.1007/s10936-004-2665-7

Holmes, V. M., & Segui, J. (2006). Assigning grammatical gender during word production. *Journal of Psycholinguistic Research, 35*(1), 5–30. https://doi.org/10.1007/s10936-005-9001-8

Hopp, H. (2013). Grammatical gender in adult L2 acquisition: Relations between lexical and syntactic variability. *Second Language Research*, *29*(1), 33–56. https://doi.org/10.1177/0267658312461803

Hopp, H. (2016). Learning (not) to predict: Grammatical gender processing in second language acquisition. *Second Language Research*, *32*(2), 277–307. https://doi.org/10.1177/0267658315624960

Ibrahim, M. H. (1973). *Grammatical Gender: Its Origin and Development*. De Gruyter Mouton. https://doi.org/10.1515/9783110905397

Jackson, C. N. (2008). Proficiency Level and the Interaction of Lexical and Morphosyntactic Information During L2 Sentence Processing. *Language and Learning*, *58*, 875–909. https://doi.org/10.1111/j.1467-9922.2008.00481.x

USC UNIVERSIDADE DE SANTIAGO

Jackson-Maldonado, D., Thal, D., Marchman, V. A., Bates, E., & Gutierrez-Clellen, V. (1993). Early lexical development in Spanish speaking infants and toddlers. *Journal of Child Language, 20*, 523–549. https://doi.org/10.1017/S0305000900008461

Jacob, R., & Laurin, J. (1994). *Ma grammaire* [My grammar]. Les Éditions françaises.

Jacobsen, T. (1999). Effects of grammatical gender on picture and word naming: evidence from German. *Journal of Psycholinguistic Research, 28*(5), 499–514. https://doi.org/10.1023/a:1023268310519

Janssen, N., & Caramazza, A. (2003). The selection of closed-class words in noun phrase production: The case of Dutch determiners. *Journal of Memory and Language, 48*(3), 635–652. https://doi.org/10.1016/s0749-596x(02)00531-4

Jeffreys, H. (1961). *Theory of probability (3rd ed.)*. Oxford University Press.

Jescheniak, J. D. (1999). Gender priming in picture naming: Modality and baseline effects. *Journal of Psycholinguistic Research, 28*(6), 729–737. https://doi.org/10.1023/A:1023229329967

Jescheniak, J. D., & Schriefers, H. (2001). Priming effects from phonologically related distractors in picture–word interference. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 54A*(2), 371–382. https://doi.org/10.1080/02724980042000273

Jescheniak, J. D., Schriefers, H., & Lemhöfer, K. (2014). Selection of freestanding and bound gender-marking morphemes in speech production: a review. *Language, Cognition and Neuroscience, 29*(6), 684–694. https://doi.org/10.1080/01690965.2012.654645

Jusczyk, P., Luce, P., & Charles-Luce, J. (1994). Infants' sensitivity to phonotactic patterns in the native language. *Journal of Memory and Language, 33*, 630–645. https://doi.org/10.1006/jmla.1994.1030

Karmiloff-Smith, A. (1979). *A Functional Approach to Child Language: A Study of Determiners and Reference*. Cambridge University Press.

Karmiloff-Smith, A., Grant, J., Berthoud, I., Davies, M., Howlin, P., & Udwin, O. (1997). Language and Williams syndrome: How intact is "intact"? *Child Development, 68*(2), 246–262. https://doi.org/10.2307/1131848

Kempen, G., & Hoenkamp, E. (1987). An incremental procedural grammar for sentence formulation. *Cognitive Science, 11*(2), 201–258. https://doi.org/10.1207/s15516709cog1102_5

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Kempen, G., & Huijbers, P. (1983). The lexicalization process in sentence production and naming: indirect election of words, *Cognition*, *14*(2), 185–209. https://doi.org/10.1016/0010-0277(83)90029-X

Keuleers, E., Brysbaert, M., & New, B. (2010). SUBTLEX-NL: A new measure for Dutch word frequency based on film subtitles. *Behavior Research Methods*, *42*, 643–650. https://doi.org/10.3758/BRM.42.3.643

Kibort, A. (2008). *Grammatical Features Inventory: Inflectional class*. University of Surrey. https://doi.org/10.15126/SMG.18/1.12

Kibort, A., & Corbett, G. G. (2008). *Grammatical Features Inventory: Gender*. University of Surrey. https://doi.org/10.15126/SMG.18/1.01

Kidd, E., Donnelly, S., & Christiansen, M. H. (2018). Individual Differences in Language Acquisition and Processing. *Trends in Cognitive Sciences*, *22*(2), 154–169. https://doi.org/10.1016/j.tics.2017.11.006

Kilarski, M. (2005). On grammatical gender as an arbitrary and redundant category. In D. A. Kibbee (Ed.), *History of Linguistics 2005: Selected papers from the Tenth International Conference on the History of the Language Sciences (ICHOLS X), 1–5 September 2005, Urbana-Champaign, Illinois* (pp. 24–36). John Benjamins Publishing Company. https://doi.org/10.1075/sihols.112.03kil

Kinoshita, S., & Verdonschot, R. G. (2020). Phonological encoding is free from orthographic influence: evidence from a picture variant of the phonological Stroop task. *Psychological Research*, *85*(3), 1340–1347. https://doi.org/10.1007/s00426-020-01315-2

Klassen, R. (2016). The representation of asymmetric grammatical gender systems in the bilingual mental lexicon. *Probus*, *28*(1). https://doi.org/10.1515/probus-2016-0002

Klein, G. S. (1964). Semantic power measured through the interference of words with color-naming. *American Journal of Psychology*, *77*, 576–588. https://doi.org/10.2307/1420768

Köpcke, K.-M. (1982). *Untersuchungen zum Genussystem der deutschen Gegenwartssprache* [Investigations on the gender system of the contemporary German language]. Tübingen.

Köpcke, K.-M., & Zubin, D. (1983). Die kognitive Organisation der Genuszuweisung zu den einsilbigen Nomen der deutschen Gegenwartssprache [The cognitive organization of

gender allocation to monosyllabic nouns in contemporary German]. *Zeitschrift für germanistische Linguistik*, *11*, 166–182.

Köpcke, K.-M., & Zubin, D. (1984). Sechs Prinzipien für die Genuszuweisung im Deutschen: Ein Beitrag zur natürlichen Klassifikation [Six principles for gender allocation in German: A Contribution to Natural Classification]. *Linguistische Berichte*, *93*, 26–50.

Korvorst, M., Roelofs, A., & Levelt, W. J. (2006). Incrementality in naming and reading complex numerals: evidence from eyetracking. *Quarterly Journal of Experimental Psychology*, *59*(2), 296–311. https://doi.org/10.1080/17470210500151691

Kramer, R. (2015). *The Morphosyntax of Gender*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199679935.001.0001

Kroll, J. F., Michael, E., Tokowicz, N., & Dufour, R. (2002). The development of lexical fluency in a second language. *Second Language Research*, *18*(2), 137–171. https://doi.org/10.1191/0267658302sr201oa

Kroll, J. F., & Stewart, E. (1994). Category Interference in Translation and Picture Naming: Evidence for Asymmetric Connections Between Bilingual Memory Representations. *Journal of Memory and Language*, *33*(2), 149–174. https://doi.org/10.1006/jmla.1994.1008

Kroll, J. F., & Tokowicz, N. (2005). Models of bilingual representation and processing: Looking back and to the future. In J. F. Kroll & A. M. B. de Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 531–553). Oxford University Press.

Kung, S. (2007). Numeral Classifiers in Lhiimaqalhqama'. In F. Hoyt, N. Seifert, A. Teodorescu & J. White (Eds.), *Texas Linguistics Society 9: Morphosyntax of Underrepresented Languages* (pp. 185–201). CSLI Publications.

Kupisch, T., Geiß, M., Mitrofanova, N., & Westergaard, M. (2018*). Gender Cues in L1 Russian Children Acquiring German as an Early L2*. EuroSLA 28 - Universität Münster. Münster, Germany.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2014). *lmerTest: Tests for random and fixed effects for linear mixed effect models: R package version 2.0-6*. http://CRAN.R-project.org/package=lmerTest

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

La Heij, W., Mak, P., Sander, J., & Willeboorsde, E. (1998).The gender-congruency effect in picture word task. *Psychological Research*, *61*(3), 209–219. https://doi.org/10.1007/s004260050026

Lakens D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, *4*, 863. https://doi.org/10.3389/fpsyg.2013.00863

Lapointe, S. G., & Dell, G. S. (1989). A synthesis of some recent work in sentence production. In G. N. Carlson & M. K. Tanenhaus (Eds.), *Linguistic structure in language processing* (pp. 107–156). Kluwer Academic Publishers.

Lass, R. (1999). Phonology and morphology. In N. Blake (Ed.), *The Cambridge History of the English Language. Vol. II (1066-1476)* (pp. 23–155). Cambridge University Press.

Leding, J. K. (2018). Adaptive memory: Animacy, threat, and attention in free recall. *Memory & Cognition*, *47*(3), 383–394. https://doi.org/10.3758/s13421-018-0873-x

Lemhöfer, K., Schriefers, H., & Jescheniak, J. D. (2006). The processing of free and bound gender-marked morphemes in speech production: Evidence from Dutch. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(2), 437–442. https://doi.org/10.1037/0278-7393.32.2.437

Lemhöfer, K., Spalek, K., & Schriefers, H. (2008). Cross-language effects of grammatical gender in bilingual word recognition and production. *Journal of Memory and Language*, *59*(3), 312–330. https://doi.org/10.1016/j.jml.2008.06.005

Leonard, L., Bortolini, U., Caselli, M. C., McGregor, K., & Sabbadini, L. (1992). Morphological deficits in children with specific language impairment: the status of features in the underlying grammar. *Language Acquisition*, *2*(2), 151–179. https://doi.org/10.1207/s15327817la0202_2

Leonard, L., Bortolini, U., Caselli, M. C., & Sabbadini, L. (1993).The use of articles by Italian-speaking children with specific language impairment. *Clinical Linguistics and Phonetics*, *7*(1), 19–27. https://doi.org/10.3109/02699209308985541

Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition*, *14*(1), 41–104. https://doi.org/10.1016/0010-0277(83)90026-4

Levelt, W. J. M. (1989). *Speaking. From intention to articulation*. The MIT Press.

USC UNIVERSIDADE DE SANTIAGO

Levelt, W. J. M. (1992). Accessing words in speech production: stages, processes and representations. *Cognition*, *42*(1-3), 1–22. https://doi.org/10.1016/0010-0277(92)90038-j

Levelt, W. J. M. (2001). Spoken word production: A theory of lexical access. *Proceedings of the National Academy of Sciences*, *98*(23), 13464–13471. https://doi.org/10.1073/pnas.231459498

Levelt, W. J. M., & Meyer, A. (2000). Word for word: Multiple lexical access in speech production. *European Journal of Cognitive Psychology*, *12*(4), 433–452. https://doi.org/10.1080/095414400750050178

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. https://doi.org/10.1017/S0140525X99001776

Levelt, W. J. M., & Schriefers, H. (1987). Stages of lexical access. In G. Kempen (Ed.), *Natural language generation: New results in artificial intelligence, psychology and linguistics*. Martinus Nijhoff.

Li, P., Zhang, F., Yu, A., & Zhao, X. (2020). Language History Questionnaire (LHQ3): An enhanced tool for assessing multilingual experience. *Bilingualism: Language and Cognition*, *23*(5), 938–944. https://doi.org/10.1017/S1366728918001153

Lim, J. H., & Christianson, K. (2013). Integrating meaning and structure in L1–L2 and L2–L1 translations. *Second Language Research*, *2*, 233–256. https://doi.org/10.1177/0267658312462019

Lim, J. H., & Christianson, K. (2015). Second language sensitivity to agreement errors: Evidence from eye movements during comprehension and translation. *Applied Psycholinguistics*, *36*(6), 1283–1315. https://doi.org/10.1017/S0142716414000290

Linck, J. A., Hoshino, N., & Kroll, J. F. (2008). Cross-language lexical processes and inhibitory control. *The mental lexicon*, *3*(3), 349–374. https://doi.org/10.1075/ml.3.3.06lin

Linck, J. A., Kroll, J. F., & Sunderman, G. (2009). Losing access to the native language while immersed in a second language: Evidence for the role of inhibition in second-language learning. *Psychological Science*, *20*(12), 1507–1515. https://doi.org/10.1111/j.1467-9280.2009.02480.x

USC UNIVERSIDADE DE SANTIAGO

Lindsey, B. A., & Gerken, L. (2012). The role of morphophonological regularity in young Spanish-speaking children's production of gendered noun phrases. *Journal of Child Language*, *39*(4), 753–776. https://doi.org/10.1017/S0305000911000250

Lipp, O. V., Derakshan, N., Waters, A. M., & Logies, S. (2004). Snakes and cats in the flower bed: Fast detection is not specific to pictures of fear-relevant animals. *Emotion*, *4*(3), 233–250. https://doi.org/10.1037/1528-3542.4.3.233

Longoni, F., Grande, M., Hendrich, V., Kastrau, F., & Huber, W. (2005). An fMRI study on conceptual, grammatical, and morpho-phonological processing. *Brain and Cognition*, *57*(2), 131–134. https://doi.org/10.1016/j.bandc.2004.08.032

Lorenz, A., & Zwitserlood, P. (2016). Semantically Transparent and Opaque Compounds in German Noun-Phrase Production: Evidence for Morphemes in Speaking. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.01943

Luce, R. D. (1959). *Individual Choice Behavior*. Wiley.

Lupker, S. J. (1979). The semantic nature of response competition in the picture-word interference task. *Memory & Cognition*, *7*(6), 485–495. https://doi.org/10.3758/BF03198265

Lupker, S. J. (1982). The role of phonetic and orthographic similarity in picture–word interference. *Canadian Journal of Psychology/Revue canadienne de psychologie*, *36*(3), 349–367. https://doi.org/10.1037/h0080652

Luraghi, S. (2011). The origin of the Proto-Indo-European gender system: Typological considerations. *Folia Linguistica*, *45*(2), 435–463. https://doi.org/10.1515/flin.2011.016

Lyster, R. (2006). Predictability in French gender attribution: A corpus analysis. *Journal of French Language Studies*, *16*(1), 69–92. https://doi.org/10.1017/S0959269506002304

MacWhinney, B. (2004). Parameters or cues? *Bilingualism: Language and Cognition*, *7*(1), 35–36. https://doi.org/10.1017/S1366728904001233

MacWhinney, B. (1987). The Competition model. In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 249–308). Lawrence Erlbaum Associates, Inc.

MacWhinney, B., Leinbach, J., Taraban, R., & McDonald, J. (1989). Language learning: Cues or rules? *Journal of Memory and Language*, *28*, 255–277 https://doi.org/10.1016/0749-596X(89)90033-8

USC UNIVERSIDADE DE SANTIAGO

Manolescu, A., & Jarema, G. (2015). Grammatical gender in Romanian-French bilinguals. *The Mental Lexicon*, *10*, 390–412. https://doi.org/10.1075/ml.10.3.04man

Marangolo, P., Piras, F., Galati, G., & Burani, C. (2006). Functional anatomy of derivational morphology. *Cortex*, *42*(8), 1093–1106. https://doi.org/10.1016/s0010-9452(08)70221-1

Maratsos, M. P. (1979). Learning how and when to use pronouns and determiners: Studies in first language development. In P. Fletcher, & M. Garman (Eds.), *Language acquisition: Studies in first language development*. Cambridge University Press.

Mariscal, S. (2008). Early acquisition of gender agreement in the Spanish noun phrase: starting small. *Journal of Child Language*, *35*, 1–29. https://doi.org/10.1017/S0305000908008908

Matthews, C. (2010). On the nature of phonological cues in the acquisition of French gender categories: Evidence from instance-based learning models. *Lingua*, *120*, 879–900. https://doi.org/10.1016/J.LINGUA.2009.06.007

Matthews, P. H. (1972). *Inflectional morphology; a theoretical study based on aspects of latin verb conjugation*. Cambridge University Press.

Matthews, P. H. (1974). *Morphology. An introduction to the theory of word-structure*. Cambridge University Press.

McGuire, P. K., Silbersweig, D. A., & Frith, C. D. (1996). Functional neuroanatomy of verbal self-monitoring. *Brain: a journal of neurology*, *119*(Pt 3), 907–917. https://doi.org/10.1093/brain/119.3.907

Mel'cuk, I. A. (1958). Statistics and the relationship between gender of French nouns and their endings. In V. J. Rozencvejg (Ed.), *Essays on Lexical Semantics* I (pp. 11–42). Skriptor.

Meyer, A. S. (2016). Same, different, or closely related: What is the relationship between language production and comprehension? *Journal of Memory and Language*, *89*, 1–7. https://doi.org/10.1016/j.jml.2016.03.002

Meyer, A. S., & Schriefers, H. (1991). Phonological facilitation in picture-word interference experiments: Effects of stimulus onset asynchrony and types of interfering stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*(6), 1146–1160. https://doi.org/10.1037/0278-7393.17.6.1146

USC UNIVERSIDADE DE SANTIAGO

Miceli, G., Turriziani, P., Caltagirone, C., Capasso, R., Tomaiuolo, F., & Caramazza, A. (2002). The neural correlates of grammatical gender: an fMRI investigation. *Journal of Cognitive Neuroscience*, *14*(4), 618–628. https://doi.org/10.1162/08989290260045855

Mills, A. E. (1986). *The acquisition of gender: A study of English and German*. Springer-Verlag. Retrieved from https://pure.uva.nl/ws/files/1894686/27859_TheAcquisitionofGender.pdf

Miozzo, M., & Caramazza, A. (1997a). On knowing the auxiliary of a verb that cannot be named: evidence for the independence of grammatical and phonological aspects of lexical knowledge. *Journal of Cognitive Neuroscience*, *9*(1), 160–166. https://doi.org/10.1162/jocn.1997.9.1.160

Miozzo, M., & Caramazza, A. (1997b). Retrieval of lexical-syntactic features in tip-of-the-tongue states. Journal of experimental psychology. *Learning, Memory, and Cognition*, *23*(6), 1410–1423. https://doi.org/10.1037//0278-7393.23.6.1410

Miozzo, M., & Caramazza, A. (1999). The selection of determiners in noun phrase production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(4), 907–922. https://doi.org/10.1016/S0749-596X(02)00531-4

Miozzo, M., Costa, A., & Caramazza, A. (2002). The absence of a gender congruency effect in romance languages: A matter of stimulus onset asynchrony? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(2), 388–391. https://doi.org/10.1037/0278-7393.28.2.388

Moore, S. (1921). Grammatical and Natural Gender in Middle English. *PMLA*, *36*(1), 79–103. https://doi.org/10.2307/457264

Morales, L., Paolieri, D., & Bajo, M. T. (2011). Grammatical gender inhibition in bilinguals. *Frontiers in Psychology*, *2*. https://doi.org/10.3389/fpsyg.2011.00284

Morales, L., Paolieri, D., Cubelli, R., & Bajo, M. T. (2014). Transfer of Spanish grammatical gender to English: Evidence from immersed and non-immersed bilinguals. *Bilingualism: Language and Cognition*, *17*(4), 700–708. https://doi.org/10.1017/S1366728914000017

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: Language and Cognition*, *19*(2), 294–310. https://doi.org/10.1017/s1366728915000176

USC UNIVERSIDADE DE SANTIAGO

Morey, R. D., & Rouder, J. N. (2015). {BAYESFACTOR}: *Computation of Bayes Factors for Common Designs. R package version 0.9.12-2*. Available online at: https://cran.r-project.org/web/packages/BayesFactor/

Morsella, E., & Miozzo, M. (2002). Evidence for a cascade model of lexical access in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(3), 555–563. https://doi.org/10.1037/0278-7393.28.3.555

Mousikou, P., Coltheart, M., Finkbeiner, M., & Saunders, S. (2010). Can the dual–route cascaded computational model of reading offer a valid account of the masked onset priming effect? *Quarterly Journal of Experimental Psychology*, *63*(5), 984–1003. https://doi.org/10.1080/17470210903156586

Muller-Gass, A., Gonthier, I., Desrochers, A., & Campbell, K. B. (2000). Multiple P3 evidence of a two-stage process in word gender decision. *Neuroreport*, *11*(16), 3527–3531. https://doi.org/10.1097/00001756-200011090-00025

Murillo, J. (1999). La marcación del género en los sustantivos del español: entre la flexión y la derivación. *Filología y Lingüística*, *25*(1), 181–192.

Naughton, J. (2005). *Czech: An Essential Grammar.* Routledge: Taylor & Francis Group.

Navarrete, E., & Costa, A. (2005). Phonological activation of ignored pictures: Further evidence for a cascade model of lexical access. *Journal of Memory and Language*, *53*(3), 359–377. https://doi.org/10.1016/j.jml.2005.05.001

New, B., Brysbaert, M., Veronis, J., & Pallier, C. (2007). The use of film subtitles to estimate word frequencies. *Applied Psycholinguistics*, *28*, 661–677. https://doi.org/10.1017/S014271640707035X

New, J. J., Cosmides, L., & Tooby, J. (2007). Category-specific attention for animals reflects ancestral priorities, not expertise. *Proceedings of the National Academy of Sciences*, *104*(42), 16598–16603. https://doi.org/10.1073/pnas.0703913104

New, J. J., Schultz, R. T., Wolf, J., Niehaus, J. L., Klin, A., German, T. C., & Scholl, B. J. (2010). The scope of social attention deficits in autism: Prioritized orienting to people and animals in static natural scenes. *Neuropsychologia*, *48*(1), 51–59. https://doi.org/10.1016/j.neuropsychologia.2009.08.008

Nichols, J. (1992). *Linguistic Diversity in Space and Time*. University of Chicago Press.

of Ockham, W. (1980). Quodlibeta septem. *Guillelmi de Ockham Opera theologica*, *9*. Bonaventure University, Franciscan Institute.

USC UNIVERSIDADE DE SANTIAGO

Ogneva, A. (2023). A review on grammatical gender acquisition in monolingual Spanish-speaking children. *Onomázein*, *59*. https://doi.org/10.7764/onomazein.59.03

Open Science Collaboration (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), 1–8. https://doi.org/10.1126/science.aac4716

O'Rourke, P. L. (2007). The gender congruency effect in bare noun production in Spanish. *Coyote Papers: Working Papers in Linguistics, Linguistic Theory at the University of Arizona*, *15*, 66–89.

O'Rourke, P. L. (2009). *The nature of syntactic gender processing in spanish: An ERP study* [Doctoral dissertation, University of Arizona]. Available from APA PsycInfo. Retrieved from https://search.proquest.com/docview/622059047?accountid=17253

Oswald, F. L., Mitchell, G., Blanton, H., Jaccard, J., & Tetlock, P. E. (2013). Predicting ethnic and racial discrimination: A meta-analysis of IAT criterion studies. *Journal of Personality and Social Psychology*, *105*(2), 171–192. https://doi.org/10.1037/a0032734

Padovani, R., Calandra-Buonaura, G., Cacciari, C., Benuzzi, F., & Nichelli, P. (2005). Grammatical gender in the brain: evidence from an fMRI study on Italian. *Brain research bulletin*, *65*(4), 301–308. https://doi.org/10.1016/j.brainresbull.2004.11.025

Palancar, E. L. (2014). A mixed system of agreement in the suffix classes of Lealao Chinantec. *Morphology*, *25*, 29–62. https://doi.org/10.1007/s11525-014-9252-x

Paolieri, D., Cubelli, R., Macizo, P., Bajo, M. T., Lotto, L., & Job, R. (2010). Grammatical gender processing in Italian and Spanish bilinguals, *The Quarterly Journal of Experimental Psychology*, *63*, 1631–1645. https://doi.org/10.1080/17470210903511210

Paolieri, D., Demestre, J., Guasch, M., Bajo, M. T., & Ferré, P. (2020). The gender congruency effect in Catalan–Spanish bilinguals: Behavioral and electrophysiological evidence. *Bilingualism: Language and Cognition*, *23*(5), 1045–1055. https://doi.org/10.1017/S1366728920000073

Paolieri, D., Lotto, L., Leoncini, D., Cubelli, R., & Job, R. (2011). Differential effects of grammatical gender and gender inflection in bare noun production. *The British Psychological Society*, *102*(1), 19–36. https://doi.org/10.1342/000712610X496536

Paolieri, D., Lotto, L., Morales, L., Bajo, M. T., Cubelli, R., & Job, R. (2010). Grammatical gender processing in romance languages: Evidence from bare noun production in Italian

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

and Spanish. *European Journal of Cognitive Psychology, 22*(3), 335–347. https://doi.org/10.1080/09541440902916803

Paolieri, D., Padilla, F., Koreneva, O., Morales, L., & Macizo, P. (2019). Gender congruency effects in Russian–Spanish and Italian–Spanish bilinguals: The role of language proximity and concreteness of words. *Bilingualism: Language and Cognition*, *22*, 112–129. https://doi.org/10.1017/s1366728917000591

Pérez-Pereira, M. (1991). The acquisition of gender: What Spanish children tell us. *Journal of Child Language*, *18*(3), 571–590. https://doi.org/10.1017/s0305000900011259

Perona-Jara, L. (2015). *Desarrollo gramatical en niños con trastorno específico del lenguaje: el género gramatical* [Doctoral dissertation, Universidad Autónoma de Madrid]. Retrieved from https://repositorio.uam.es/bitstream/handle/10486/671493/perona_jara_laura.pdf?sequence=1

Perret, C., & Bonin, P. (2019). Which variables should be controlled for to investigate picture naming in adults? A Bayesian meta-analysis. *Behavior Research Methods*, *51*, 2533–2545. https://doi.org/10.3758/s13428-018-1100-1

Peterson, R. R., & Savoy, P. (1998). Lexical selection and phonological encoding during language production: Evidence for cascaded processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(3), 539–557. https://doi.org/10.1037/0278-7393.24.3.539

Pinheiro, J. C., & Bates D. M. (2000). *Mixed-Effects Models in S and S-PLUS*. Springer-Verlag.

Pivneva, I., Palmer, C., & Titone, D. (2012). Inhibitory control and l2 proficiency modulate bilingual language production: evidence from spontaneous monologue and dialogue speech. *Frontiers in psychology*, *3*, 3–57. https://doi.org/10.3389/fpsyg.2012.00057

Porto Editora (2003-2021). *Dicionário Porto Editora Infopédia* [Infopédia Dictionary from Porto Editora]. Available at https://www.infopedia.pt/

Postiglione, F., De Martino, M., Bracco, G., & Laudanna, A. (2014). *Frequency distribution of inflectional classes and grammatical gender: a lexical database for Italian nouns* [Poster presentation]. Secondo Convegno Interannuale della Società di Linguistica Italiana (SLI): Teorie e approcci usage-based in linguistica, Bolzano, Italy.

Prado, M. (1982). El género en español y la teoría de la marcadez [Gender in Spanish and the markedness theory]. *Hispania*, *65*(2), 258–266. https://doi.org/10.2307/341541

Priberam (2008-2021). *Dicionário Priberam da Língua Portuguesa* [Priberam Dictionary of the Portuguese Language]. Available at https://dicionario.priberam.org

Prior, A., MacWhinney, B., & Kroll, J. F. (2007). Translation norms for English and Spanish: The role of lexical variables, word class, and L2 proficiency in negotiating translation ambiguity. *Behavior Research Methods*, *39*, 1029–1038. https://doi.org/10.3758/BF03193001

Protopapas, A. (2007). Check Vocal: A program to facilitate checking the accuracy and response time of vocal responses from DMDX. *Behavior Research Methods*, *39*(4), 859–862. https://doi.org/10.3758/bf03192979

Qinfang, Z., & Yufang, Y. (2003). The determiners of picture-naming latency. *Acta Psychologica Sinica*, *35*(04), 447–454.

Quiñones, I., Molinaro, N., Mancini, S., Hernández-Cabrera, J. A., Barber, H., & Carreiras, M. (2018). Tracing the interplay between syntactic and lexical features: fMRI evidence from agreement comprehension. *NeuroImage*, *175*, 259–271. https://doi.org/10.1016/j.neuroimage.2018.03.069

R Core Team (2018). *R: A language and environment for statistical computing. R Foundation for Statistical Computing*, Vienna, Austria. URL: https://www.R-project.org/

Radanović, J., Westbury, C., & Milin, P. (2016). Quantifying semantic animacy: How much are words alive? *Applied Psycholinguistics*, *37*(6), 1477–1499. https://doi.org/10.1017/s0142716416000096

Radeau, M., Mousty, P., & Bertelson, P. (1989). The effect of the uniqueness point in spoken-word recognition. *Psychological Research*, *51*, 123–128. https://doi.org/10.1007/BF00309307

Regula, C. R. (1971). Some suggestions for improving the psychology laboratory course experience. *American Psychologist*, *26*(11), 1020–1021. https://doi.org/10.1037/h0032444

Rio-Torto, G., Soares-Rodrigues, A., Pereira, I., Pereira, R., & Ribeiro, S. (2013). *Gramática derivacional do Português* [Derivational grammar of Portuguese]. Imprensa da Universidade de Coimbra. https://doi.org/10.14195/978-989-26-0864-

Ritchie, S. (2020). *Science Fictions: Exposing Fraud, Bias, Negligence, and Hype in Science*. The Bodley Head.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition*, *42*(1-3), 107–142. https://doi.org/10.1016/0010-0277(92)90041-F

Roelefs, A. (1993). Testing a non-decompositional theory of lemma retrieval in speaking: Retrieval of verbs. *Cognition*, *47*(1), 59–87. https://doi.org/10.1016/0010-0277(93)90062-Z

Roelofs, A. (1996a). Morpheme frequency in speech production: testing WEAVER. In G. E. Booij & J. van Marle (Eds.), *Yearbook of Morphology* (pp. 135–154). Kluwer Academic.

Roelofs, A. (1996b). Serial order in planning the production of successive morphemes of a word. *Journal of Memory and Language*, *35*(6), 854–876. https://doi.org/10.1006/jmla.1996.0044

Roelofs, A. (1997). The WEAVER model of word-form encoding in speech production. *Cognition*, *64*(3), 249–284. https://doi.org/10.1016/S0010-0277(97)00027-9

Roelofs, A. (2004). Comprehension-Based Versus Production-Internal Feedback in Planning Spoken Words: A Rejoinder to Rapp and Goldrick (2004). *Psychological Review*, *111*(2), 579–580. https://doi.org/10.1037/0033-295X.111.2.579

Roelofs, A., Meyer, A. S., & Levelt, W. J. (1998). A case for the lemma/lexeme distinction in models of speaking: comment on Caramazza and Miozzo (1997). *Cognition*, *69*(2), 219–230. https://doi.org/10.1016/s0010-0277(98)00056-0

Roelofs, A. (2003). Goal-referenced selection of verbal action: Modeling attentional control in the Stroop task. *Psychological Review*, *110*(1), 88–125. https://doi.org/10.1037/0033-295X.110.1.88

Roelofs, A., & Piai, V. (2015). Aspects of competition in word production: Reply to Mahon and Navarrete. *Cortex: A Journal Devoted to the Study of the Nervous System and Behavior*, *64*, 420–424. https://doi.org/10.1016/j.cortex.2014.10.016

Rosinski, R. R., Golinkoff, R. M., & Kukish, K. S. (1975). Automatic semantic processing in a picture-word interference task. *Child Development*, *48*, 643–647. https://doi.org/10.2307/1128859

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16*(2), 225–237. https://doi.org/10.3758/PBR.16.2.225

USC UNIVERSIDADE DE SANTIAGO

Russo, A. G., Esposito, F., Laudanna, A., Mancuso, A., Di Salle, F., Elia, A., & De Martino, M. (2021). The neural substrate of noun morphological inflection: A rapid event-related fMRI study in Italian. *Neuropsychologia*, *151*, 107–699. https://doi.org/10.1016/j.neuropsychologia.2020.107699

Sá-Leite, A. R., Fraga, I., & Comesaña, M. (2019). Grammatical gender processing in bilinguals: An analytic review. *Psychonomic Bulletin & Review*, *26*(4), 1148–1173. https://doi.org/10.3758/s13423-019-01596-8

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

Sagarra, N., & Herschensohn, J. (2011). Proficiency and animacy effects on L2 gender agreement processes during comprehension. *Language Learning*, *61*(1), 80–116. https://doi.org/10.1111/j.1467-9922.2010.00588.x

Said Ali, M. (1921). *Gramatica histórica da língua portuguesa* [Historical grammar of the Portuguese language]. Edições Melhoramento.

Salamoura, A., & Williams, J. N. (2007). The representation of grammatical gender in the bilingual lexicon: Evidence from Greek and German. *Bilingualism: Language and Cognition*, *10*, 257–275. https://doi.org/10.1017/s1366728907003069

Schiller, N. O. (2004). The onset effect in word naming. *Journal of Memory and Language*, *50*(4), 477–490. https://doi.org/10.1016/j.jml.2004.02.004

Schiller, N. O., Münte, T. F., Horemans, I., & Jansma, B. M. (2003). The influence of semantic and phonological factors on syntactic decisions: an event-related brain potential study. *Psychophysiology*, *40*(6), 869–877. https://doi.org/10.1111/1469-8986.00105

Schiller, N. O., & Caramazza, A. (2003). Grammatical feature selection in noun phrase production: Evidence from German and Dutch. *Journal of Memory and Language*, *48*(1), 169–194. https://doi.org/10.1016/S0749-596X(02)00508-9

Schiller, N. O., & Caramazza, A. (2006). Grammatical gender selection and the representation of morphemes: The production of Dutch diminutives. *Language and Cognitive Processes*, *21*(7-8), 945–973. https://doi.org/10.1080/01690960600824344

Schiller, N. O., & Costa, A. (2006). Different selection principles of freestanding and bound morphemes in language production. *Journal of Experimental Psychology: Learning,*

Memory, and Cognition, 32(5), 1201–1207. https://doi.org/10.1037/0278-7393.32.5.1201

Schmidt, R. (1986). Was weiss der Artikel vom Hauptwort? Em Beitrag zur Verarbeitung syntaktischer Beziehungen beim Lesen [What does the article know about the noun? Contribution to the processing of syntactic relationships when reading]. *Zeitschrift fur experimentelle and angewandte Psychologie, 33*, 150–163.

Schriefers, H. (1993). Syntactic processes in the production of noun phrases. *Journal of Experimental Psychology: Learning, memory, and Cognition, 19*(4), 841–850. https://doi.org/10.1037/0278-7393.19.4.841

Schriefers, H., Jescheniak, J. D., & Hantsch, A. (2002). Determiner selection in noun phrase production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(5), 941–950. https://doi.org/10.1037/0278-7393.28.5.941

Schriefers, H., Jescheniak, J. D., & Hantsch, A. (2005). Selection of gender-marked morphemes in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(1), 159–168. https://doi.org/10.1037/0278-7393.31.1.159

Schriefers, H., & Teruel, E. (2000). Grammatical gender in noun phrase production: The gender interference effect in German. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(6), 1368–1377. https://doi.org/10.1037/0278-7393.26.6.1368

Schriefers, H., & Jescheniak, J. (1999). Representation and processing of grammatical gender in language production: A review. *Journal of Psycholinguistic Research, 28*(6), 575–600. https://doi.org/10.1023/A:1023264810403

Schwichtenberg, B., & Schiller, N. O. (2004). Semantic gender assignment regularities in German. *Brain and language, 90*(1-3), 326–337. https://doi.org/10.1016/S0093-934X(03)00445-0

Schwindt, L. C. (2018). Exponência de gênero e classe temática em português brasileiro [Exponence of gender and thematic class in Brazilian Portuguese]. *Delta, 34*(2), 745–768. https://doi.org/10.1590/0102-44509154975775546

Sedlmeier, P., Tipandjan, A., & Jänchen, A. (2016). How Persistent are Grammatical Gender Effects? The Case of German and Tamil. *Journal of Psycholinguistic Research, 45*(2), 317–336. https://doi.org/10.1007/s10936-015-9350-x

Segalowitz N. (2010). *The Cognitive Bases of Second Language Fluency*. Routledge.

USC UNIVERSIDADE DE SANTIAGO

Segalowitz N., & Hulstijn J. (2005). Automaticity in bilingualism and second language learning. In J. F. Kroll & A. B. B. de Groot (Eds.), *Handbook of Bilingualism: Psycholinguistic Approaches* (pp. 371–388). Oxford University Press.

Séguin, H. (1969). *Les marques du genre dans le lexique français écrit contemporain: Compilation des cas et essai de classement* [The marks of the genre in the written French lexicon contemporary: Compilation of cases and trial of classification]. Mémoire de D.E.S. Université de Montréal, Département de Linguistique.

Seigneuric, A., Zagar, D., Meunier, F., & Spinelli, E. (2007). The relation between language and cognition in 3- to 9-year-olds: the acquisition of grammatical gender in French. *Journal of Experimental Child Psychology*, *96*(3), 229–246. https://doi.org/10.1016/j.jecp.2006.12.003

Shadid, A. C. (1997). La formación del género en español. *Filología y Lingüística*, *23*(2), 153–159. Retrieved from https://revistas.ucr.ac.cr/index.php/filyling/article/download/20410/20661/0

Siemund, P. (2013). *Pronominal gender in English: A study of English varieties from a cross-linguistic perspective* (third edition). Routledge. https://doi.org/10.4324/9780203455944

Sinnemäki., K. T.-K. (2019). On the distribution and complexity of gender and numeral classifiers. In F. di Garbo, B. Olsson & B. Wälchli (Eds.), *Grammatical gender and linguistic complexity II: Volume II: World-wide comparative studies* (pp. 133–200). Language Science Press.

Soares, A. P., Costa, A. S., Machado, J., Comesaña, M., & Oliveira, H. (2017). The Minho Word Pool: Norms for imageability, concreteness and subjective frequency for 3,800 Portuguese words. *Behavior Research Methods*, *49*(3), 1065–1081. https://doi.org/10.3758/s13428-016-0767-4

Soares, A. P., Iriarte, A., Almeida, J. J., Simões, A., Costa, A., Machado, J., França, P., Comesaña, M., Rauber, A., Rato, A., & Perea, M. (2018). Procura-PALavras (P-PAL): A web-based interface for a new European Portuguese lexical database. *Behavior Research Methods*, *50*(4), 1461–1481. https://doi.org/10.3758/s13428-018-1058-z

Soares, A. P., Lages, A., Silva, A., Comesaña, M., Sousa, I., Pinheiro, A. P., & Perea, M. (2019). Psycholinguistic variables in visual word recognition and pronunciation of European

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Portuguese words: A mega-study approach. *Language, Cognition and Neuroscience*, *34*(6), 689–719. https://doi.org/10.1080/23273798.2019.1578395

Soares, A. P., Machado, J., Costa, A., Iriarte, Á., Simões, A., de Almeida, J. J., Comesaña, M., & Perea, M. (2015). On the Advantages of Word Frequency and Contextual Diversity Measures Extracted from Subtitles: The Case of Portuguese. *Quarterly Journal of Experimental Psychology*, *68*(4), 680–696. https://doi.org/10.1080/17470218.2014.964271

Soares, A. P., Medeiros, J. C., Simoes, A., Machado, J., Costa, A., Iriarte, A., & Comesaña, M. (2014). ESCOLEX: A grade-level lexical database from European Portuguese elementary to middle school textbooks. *Behavior Research Methods*, *46*, 240–253. https://doi.org/10.3758/s13428-013-0350-1

Soares, A. P., Oliveira, H., Comesaña, M., & Costa, A. (2018). Lexico-syntactic interactions in the resolution of second language relative clause ambiguities: The role of cognate status and second language proficiency. *Psicológica*, *39*, 164–197. https://doi.org/10.2478/psicolj-2018-0008

Soares, A. P., Oliveira, H., Ferreira, M., Comesaña, M., Macedo, F., Ferré, P., Acuña-Fariña, J. C., Hernández-Cabrera, J., & Fraga, I. (2019). The role of cognate status and second language proficiency on relative clause attachment preferences in a second language: An eye-tracker study. *PLOS One*.

Spalek, K., Franck, J., Schriefers, H., & Frauenfelder, U. H. (2008). Phonological regularities and grammatical gender retrieval in spoken word recognition and word production. *Journal of Psycholinguistic Research*, *37*(6), 419–442. https://doi.org/10.1007/s10936-008-9074-2

Spalek, K., & Schriefers, H. J. (2005). Dominance affects determiner selection in language production. *Journal of Memory and Language*, *52*(1), 103–119. https://doi.org/10.1016/j.jml.2004.09.001

Starreveld, P., & La Heij, W. (2004). Phonological facilitation of grammatical gender retrieval. *Language and Cognitive Processes*, *19*(6), 677–711. https://doi.org/10.1080/01690960444000061

Steele, S. (1978). Word order variation. A typological study. In J. H. Greenberg, C. A. Ferguson & E. A. Moravcsik (Eds.), *Universals of Human Language. Volume IV, Syntax* (pp. 585–623). Stanford University Press.

USC UNIVERSIDADE DE SANTIAGO

Stemberger, J. (1985). An interactive activation model of language production. In A. W. Ellis (Ed.), *Progress in the Psychology of Language, Vol. 1* (pp. 143–186). Lawrence Erlbaum Associates.

Szagun, G., Stumper, B., Sondag, N., & Franik, M. (2007). The acquisition of gender marking by young German-speaking children: Evidence for learning guided by phonological regularities. *Journal of Child Language*, *34*, 445–471. https://doi.org/10.1017/S0305000906007951

Szekely, A., D'Amico, S., Devescovi, A., Federmeier, K., Herron, D., Iyer, G., Jacobsen, T., Arévalo, A., Vargha, A., & Bates, E. (2005). Timed action and object naming. *Cortex*, *41*(1), 7–25. https://doi.org/10.1016/s0010-9452(08)70174-6

Szekely, A., Jacobsen, T., D'Amico, S., Devescovi, A., Andonova, E., Herron, D., Lu, C. C., Pechmann, T., Pléh, C., Wicha, N., Federmeier, K., Gerdjikova, I., Gutierrez, G., Hung, D., Hsu, J., Iyer, G., Kohnert, K., Mehotcheva, T., Orozco-Figueroa, A., Tzeng, A., ... & Bates, E. (2004). A new on-line resource for psycholinguistic studies. *Journal of Memory and Language*, *51*(2), 247–250. https://doi.org/10.1016/j.jml.2004.03.002

Szewczyk, J. M., & Schriefers, H. (2011). Is animacy special? *Brain Research*, *1368*, 208–221. https://doi.org/10.1016/j.brainres.2010.10.070

Taft, M., & Meunier, F. (1998). Lexical representation of gender: a quasiregular domain. *Journal of Psycholinguistic Research*, *27*(1), 23–45. https://doi.org/10.1023/a:1023270723066

Teschner, R. V., & Russell, W. M. (1984). The gender patterns of Spanish nouns: An inverse dictionary-based analysis. *Hispanic Linguistics*, *1*, 115–132.

Thal, D., Jackson-Maldonado, D., & Acosta, D. (2000). Validity of a parent-report measure of vocabulary and grammar for Spanish-speaking toddlers. *Journal of Speech, Language, and Hearing Research*, *43*, 1087–1100. https://doi.org/10.1044/jslhr.4305.1087

Tipton, E. (2015). Small sample adjustments for robust variance estimation with meta-regression. *Psychological Methods*, *20*(3), 375–393. https://doi.org/10.1037/met0000011

Trudgill, P. (1999). Language contact and the function of linguistic gender. *Poznań Studies in Contemporary Linguistics*, *35*, 133–152. Retrieved from http://ifa.amu.edu.pl/psicl/files/35/10Trudgill.pdf

USC UNIVERSIDADE DE SANTIAGO

Tucker, G. R., Lambert, W., & Rigault, A. (1977). *The French Speaker's Skill with Grammatical Gender: An Example of Rule-governed Behaviour*. Mouton. https://doi.org/10.1515/9783110805413

Tucker, G. R., Lambert, W. E., Rigault, A. A., & Segalowitz, N. (1968). A psychological investigation of French speakers' skill with grammatical gender. *Journal of Verbal Learning and Verbal Behavior*, *7*(2), 312–316. https://doi.org/10.1016/S0022-5371(68)80007-6

Turennout, M., Hagoort, P., & Brown, C. M. (1998). Brain activity during speaking: from syntax to phonology in 40 milliseconds. *Science*, *280*(5363), 572–574. https://doi.org/10.1126/science.280.5363.572

Urrutia, M., Domínguez, A., & Álvarez, C. J. (2009). Gender activation in transparent and opaque words. *Psicothema*, *21*(1), 1–8.

VanArsdall, J. E., Nairne, J. S., Pandeirada, J. N. S., & Cogdill, M. (2014). Adaptive memory: Animacy effects persist in paired-associate learning. *Memory*, *23*(5), 657–663. https://doi.org/10.1080/09658211.2014.916304

van Berkum, J. J. A. (1996). *The psycholinguistics of grammatical gender: Studies in language comprehension and production* [Doctoral dissertation, Max Planck Institute for Psycholinguistics].

van Berkum, J. J. A. (1997). Syntactic processes in speech production: The retrieval of grammatical gender. *Cognition*, *64*(2), 115–152. https://doi.org/10.1016/S0010-0277(97)00026-7

Van Hell, J. G., & Tanner, D. (2012). Second Language Proficiency and Cross-Language Lexical Activation. *Language Learning*, *62*, 148–171. https://doi.org/10.1111/j.1467-9922.2012.00710.x

van Heugten, M., & Shi, R. (2009). French-learning toddlers use gender information on determiners during word recognition. *Developmental* Science *12*(3), 419–425. https://doi.org/10.1111/j.1467-7687.2008.00788.x

Vasishth, S., & Gelman, A. (2021). How to embrace variation and accept uncertainty in linguistic and psycholinguistic data analysis. *Linguistics*. https://doi.org/10.31234/osf.io/zcf8s

Velnić, M. (2020) Acquisition of a transparent gender system: A comparison of Italian and Croatian. *Frontiers*, *11*. https://doi.org/10.3389/fpsyg.2020.571674

USC UNIVERSIDADE DE SANTIAGO

Viana, M. C., Andrade, E., Oliveira, L. C., & Trancoso, I. M. (1991). Ler P.E.: Um utensílio para o estudo da ortografía do português [Read P.E.: A resource to study the Portuguese orthography]. In Direção da Associação Portuguesa de Lingüística (Eds.), *Actas do VII Encontro da Associação Portuguesa de Linguística, Lisbon* (pp. 474–489). Colibri-Artes gráficas.

Viechtbauer, W. (2010). Conducting Meta-Analyses in R with the metafor Package. *Journal of Statistical Software*, *36*(3), 1–48. https://doi.org/10.18637/jss.v036.i03

Vigliocco, G., Antonini, T., & Garrett, M. F. (1997). Grammatical gender is on the tip of Italian tongues. *Psychological Science*, *8*(4), 314–317. https://doi.org/10.1111/j.1467-9280.1997.tb00444.x

Vigliocco, G., Lauer, M., Damian, M., & Levelt, W. J. (2002). Semantic and syntactic forces in noun phrase production. Journal of experimental psychology. *Learning, Memory, and Cognition*, *28*(1), 46–58. https://doi.org/10.1037//0278-7393.28.1.46

Vigliocco, G., Vinson, D. P., Indefrey, P., Levelt, W. J. M., & Hellwig, F. (2004). Role of Grammatical Gender and Semantics in German Word Production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(2), 483–497. https://doi.org/10.1037/0278-7393.30.2.483

Vilela, M. (1973). Considerações gerais sobre o género: O género dentro das categorias gramaticais. *Revista da Faculdade de Letras: Filologia*, *1*, 139–150. Retrieved from https://repositorio-aberto.up.pt/bitstream/10216/8190/2/1262.pdf

Villalva, A. (2008). *Morfologia do Português* [Morphology of the Portuguese language]. Universidade Aberta.

von der Gabelentz, G. (1891). *Die Sprachwissenschaft: Ihre Aufgaben, Methoden und bisherigen Ergebnisse* [Linguistics: its tasks, methods and results so far]. Tauchnitz.

von Grebmer Zu Wolfsthurn, S., Pablos, L., & Schiller, N. O. (2021). Noun-phrase production as a window to language selection: An ERP study. *Neuropsychologia*, *162*(108055). Advance online publication. https://doi.org/10.1016/j.neuropsychologia.2021.108055

Wang, M., Chen, Y., & Schiller, N. O. (2019). Lexico-syntactic features are activated but not selected in bare noun production: Electrophysiological evidence from overt picture naming. *Cortex*, *116*, 294–307. https://doi.org/10.1016/j.cortex.2018.05.014

Wang, M., & Schiller, N. O. (2019). A review on grammatical gender agreement in speech production. *Frontiers in Psychology*, *9*. https://doi.org/10.3389/fpsyg.2018.02754

USC UNIVERSIDADE DE SANTIAGO

Warden, M. (1997). *The effect of form-focused instruction on control over grammatical gender by French immersion students in grade 11* [Doctoral dissertation, University of Toronto].

Weber, A., & Paris, G. (2004). The origin of the linguistic gender effect in spoken-word recognition: Evidence from non-native listening. *Proceedings of the Twenty-Sixth Annual Meeting of the Cognitive Science Society*, 1446–1451.

Wegener, H. (1995). *Die Nominalflexion des Deutschen—verstanden als Lerngegenstand* [The nominal inflection of German—looked at as a learning matter]. Tübingen.

Williams, S., & Hammarberg, B. (2005). Language switches in L3 production: Implications for a polyglot speaking model. *Processes in Third Language Acquisition*, *19*, 28–70. https://doi.org/10.3366/edinburgh/9780748635115.003.0003

Wilshire, C., Singh, S., & Tattersall, C. (2016). Serial order in word form retrieval: New insights from the auditory picture–word interference task. *Psychonomic Bulletin & Review*, *23*(1), 299–305. https://doi.org/10.3758/s13423-015-0882-8

Xiao, X., Dong, Q., Chen, C., & Xue, G. (2016). Neural pattern similarity underlies the mnemonic advantages for living words. *Cortex*, *79*, 99–111. https://doi.org/10.1016/j.cortex.2016.03.016

Zhang, H., Carlson, M. T., & Diaz, M. T. (2020). Investigating the effects of phonological neighbours on word retrieval and phonetic variation in word naming and picture naming paradigms. *Language, Cognition and Neuroscience*, *35*(8), 980–991. https://doi.org/10.1080/23273798.2019.1686529

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

APPENDIX A: Experimental materials for Study 1

Table A1

Target and distractor nouns for Experiment 1 (first and second conditions of animacy presence)

Exp. 1 - First Condition						**Exp. 1 – Second Condition**					
Fem. Targets	**Distractors**		**Masc. Targets**	**Distractors**		**Fem. Targets**	**Distractors**		**Masc. Targets**	**Distractors**	
mesa (table)	praia	beach	carro (car)	vento	wind	mulher (woman)	análise	analysis	homem (man)	filme	movie
	arte	art		vale	valley		cultura	culture		livro	book
	palco	stage		data	date		futebol	soccer		fonte	fountain
	sol	sun		fome	hunger		diálogo	dialogue		água	water
janela (window)	letra	letter	vinho (wine)	braço	arm	vaca (cow)	corda	rope	peixe (fish)	exame	exam
	fraude	fraud		parque	park		nave	ship		evento	event
	navio	ship		praça	square		figo	fig		frase	phrase
	exame	exam		ponte	bridge		lume	fire		capela	chapel
cama (bed)	lenda	legend	termo (thermos)	alvo	target	galinha (chicken)	torneira	tap	cavalo (horse)	ofício	craft
	dor	pain		sinal	signal		semente	seed		azeite	olive oil
	lago	lake		pedra	stone		martelo	hammer		placa	board
	lume	fire		carne	flesh		detalhe	detail		neve	snow
cadeira (chair)	farinha	flour	relógio (watch)	tabaco	tobacco	águia (eagle)	cifra	digit	rato (rat)	molho	sauce
	nave	ship		certame	event		creche	nursery		iate	yacht
	senado	senate		medalha	medal		hélio	helium		faca	knife
	teste	test		árvore	tree		cabaz	basket		mente	mind
carteira (wallet)	revolta	revolt	prato (plate)	circo	circus	ovelha (sheep)	medula	marrow	dinossauro (dinosaur)	telescópio	telescope
	gravidade	gravity		iate	yacht		cabine	cabin		chocolate	chocolate
	tornado	tornado		espada	sword		soneto	sonnet		ferramenta	tool
	recital	recital		cruz	cross		arame	wire		extremidade	limb
bicicleta (bicycle)	camisola	sweater	copo (glass)	dedo	finger	cabra (goat)	túnica	robe	pássaro (bird)	ginásio	gymnasium
	tempestade	storm		poste	post		tarte	pie		vinagre	vinegar
	catálogo	catalog		horta	vegetable		astro	star		lâmpada	lamp
	passaporte	passport			garden		capote	cloak		manchete	headline
				peste	plague						

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

flor	maré	tide	globo	joelho	knee	raposa	orelha	ear	polvo	crânio	skull
(flower)	dieta	diet	(globe)	tomate	tomato	(fox)	foice	scythe	(octopus)	bosque	grove
	arroz	rice		raiva	rage		terraço	terrace		tecla	key
	ombro	shoulder		raíz	root		ciclone	cyclone		fronte	forehead
garrafa	fogueira	fire pit	queijo	sismo	earthquake	aranha	cevada	barley	crocodilo	relâmpago	lightning
(bottle)	validade	expiration	(cheese)	baile	ball, dance	(spider)	estante	bookcase	(crocodile)	detergente	detergent
	terraço	terrace		relva	grass		gelado	ice cream		azeitona	olive
	tanque	tank		elite	elite		dique	dike		tangente	tangent
gaveta	prenda	gift	casaco	moinho	mill	mosca	panela	pan	sapo	trigo	wheat
(drawer)	claque	fan group	(jacket)	apetite	appetite	(fly)	couve	cabbage	(toad)	pote	bowl
	casino	casino		barriga	belly		lábio	lip		rolha	cork
	arame	wire		praxe	hazing		truque	trick		vide	vine branch
lata	mola	clothespin	tapete	bigode	moustache	bruxa	toalha	towel	papagaio	antídoto	antidote
(can)	cave	basement	(rug)	útero	uterus	(witch)	sande	sandwich	(parrot)	sabonete	soap
	selo	seal		morgue	morgue		lenço	scarf		amêndoa	almond
	dente	tooth		galáxia	galaxy		cofre	safe box		gabardine	gabardine
gravata	cebola	onion	berço	crânio	skull	lagosta	bússola	compass	golfinho	marmelo	quince
(tie)	estante	bookcase	(cradle)	farol	lighthouse	(lobster)	falange	phalanx	(dolphin)	peluche	plush
	centeio	rye		tecla	key		charuto	cigar		amêijoa	clam
	bosque	grove		foice	scythe		abacate	avocado		esfinge	sphinx
toalha	âncora	anchor	frigorífico	microscópio	microscope	tartaruga	beterraba	beet	camelo	fungo	fungus
(towel)	couve	cabbage	(refrigerator)	transplante	transplant	(turtle)	avalanche	avalanche	(camel)	frete	freight
	átomo	atom		recompensa	reward		candeeiro	lamp		saliva	saliva
	cabaz	basket		tuberculose	tuberculosis		esparguete	spaghetti		acidez	acidity, sourness
chaminé	entorse	sprain	armário	veneno	poison	avestruz	maionese	mayonnaise	mosquito	azevinho	holly
(chimney)	lágrima	tear	(closet)	vinagre	vinegar	(ostrich)	gelatina	jelly	(mosquito)	disfarce	disguise
	foguete	rocket		lanterna	flashlight		enxame	swarm		alcatifa	carpet
	teclado	keyboard		diálise	dialysis		remorso	remorse		planície	plain
torneira	almofada	pillow	colete	ventre	abdomen	borboleta	mangueira	hose	caranguejo	termómetro	thermometer
(tap)	dinamite	dynamite	(vest)	frasco	bottle	(butterfly)	obesidade	obesity	(crab)	piquenique	picnic
	carimbo	stamp		lápide	tombstone		eucalipto	eucalyptus		trincheira	trench
	barrete	cap		fábula	fable		envelope	envelope		sanduíche	sandwich
mochila	alergia	allergy	sapato	azulejo	tile	joaninha	fotocópia	photocopy	lagarto	tempero	seasoning
(backpack)	cabine	cabin	(shoe)	cacete	baguette	(ladybug)	disquete	floppy disk	(lizard)	chicote	whip
	cotovelo	elbow		botija	cylinder		isqueiro	lighter		camélia	camellia
	ciclone	cyclon e		prótese	prosthesis		rabanete	radish		maquete	model

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

chávena	bússola	compass	aquário	tempero	seasoning	abelha	peruca	wig	veado	palito	toothpick
(cup)	sirene	siren	(aquarium)	disfarce	disguise	(bee)	hélice	propeller	(deer)	cone	cone
	isqueiro	lighter		pulseira	bracelet		estojo	case		pomada	ointment
	diamante	diamond		hélice	propeller		quiosque	kiosk		grade	grille
vassoura	esmeralda	emerald	vaso	fungo	fungus	coruja	acácia	acacia	morcego	fósforo	match
(broom)	maionese	mayonnaise	(vase)	pote	bowl	(owl)	elipse	ellipse	(bat)	cabide	hanger
	parafuso	screw		telha	roof tile		vinilo	vinyl		cereja	cherry
	fiambre	ham		trave	beam		xarope	syrup		epiderme	epidermis
rolha	laca	lacquer	candeeiro	brinquedo	toy	minhoca	peçonha	venom	hipopótamo	candelabro	chandelier
(cork)	lente	lens	(lamp)	envelope	envelope	(worm)	laringe	larynx	(hippopotamus)	espinafre	spinach
	mofo	mold		azeitona	olive		topázio	topaz		cambalhota	back flip
	cone	cone		sanduíche	sandwich		foguete	rocket		salinidade	salinity
sanita	geleia	jam	fósforo	cogumelo	mushroom	girafa	bolota	acorn	pelicano	biscoito	cookie
(toilet)	tarte	pie	(match)	chicote	whip	(giraffe)	virose	virosis	(pelicano)	xilofone	xylophone
	baloiço	swing		abóbora	pumpkin		caroço	pit		espátula	spatula
	cabide	hanger		esfinge	sphinx		colete	vest		limousine	limousine
cesta	borga	spree (fun)	cinzeiro	morango	strawberry	zebra	sarna	scabies	esquilo	repolho	cabbage
(basket)	grade	grille	(astray)	alfinete	pin	(zebra)	coice	kick	(squirrel)	barrote	beam
	feno	hay		gelatina	gelatine		beiço	lip		alheira	type of sausage
	vime	wicker		laringe	larynx		pevide	fruit seed		raquete	racket

Note. From top to bottom of the distractors columns: gender congruent nouns with the same transparency category of that of the target noun, gender congruent nouns with a different transparency category, gender incongruent nouns with the same transparency category, gender incongruent nouns with a different transparency category.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Table A2

Target and distractor nouns for Experiment 1 (third condition of animacy presence)

Feminine Targets	**Distractors**		**Masculine Targets**	**Distractors**	
bicicleta	camisola	sweater	peixe	exame	exam
(bicycle)	tempestade	storm	(fish)	evento	event
	catálogo	catalog		frase	phrase
	passaporte	passport		capela	chapel
vaca	corda	rope	globo	joelho	knee
(cow)	nave	ship	(globe)	tomate	tomato
	figo	fig		raiva	rage
	lume	fire		raíz	root
toalha	âncora	anchor	cavalo	ofício	craft
(towel)	couve	cabbage	(horse)	azeite	olive oil
	átomo	atom		placa	board
	cabaz	basket		neve	snow
águia	cifra	digit	tapete	bigode	moustache
(eagle)	creche	nursery	(rug)	útero	uterus
	hélio	helium		morgue	morgue
	duche	basket		galáxia	galaxy
chaminé	entorse	sprain	polvo	bagaço	skull
(chimney)	lágrima	tear	(octopus)	colete	vest
	foguete	rocket		fralda	diaper
	teclado	keyboard		fronte	forehead
aranha	cevada	barley	berço	crânio	skull
(spider)	estante	bookcase	(cradle)	farol	lighthouse
	gelado	ice cream		tecla	key
	dique	dike		coice	kick
torneira	almofada	pillow	crocodilo	relâmpago	lightning
(tap)	dinamite	dynamite	(crocodile)	detergente	detergent
	carimbo	stamp		gabardina	raincoat
	barrete	cap		tangente	tangent
mosca	panela	pan	frigorífico	microscópio	microscope
(fly)	sande	sandwich	(refrigerator)	transplante	transplant
	lábio	scarf		recompensa	reward
	truque	safe box		tuberculose	tuberculosis
mochila	alergia	allergy	golfinho	marmelo	quince
(backpack)	cabine	cabin	(dolphin)	peluche	plush
	cotovelo	elbow		castanha	chestnut
	ciclone	cyclone		esfinge	sphinx
tartaruga	beterraba	beet	sapato	azulejo	tile
(turtle)	avalanche	avalanche	(shoe)	cacete	baguette
	guardanapo	napkin		botija	cylinder
	esparguete	spaghetti		prótese	prosthesis
vassoura	esmeralda	emerald	camelo	fungo	fungus
(broom)	cicatriz	scar	(camel)	frete	freight
	parafuso	screw		saliva	saliva
	fiambre	ham		acidez	acidity, sourness
avestruz	maionese	mayonnaise	candeeiro	brinquedo	toy

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

(ostrich)	gelatina	jelly	(lamp)	envelope	envelope
	enxame	swarm		azeitona	olive
	remorso	remorse		sanduíche	sandwich
rolha	laca	lacquer	morcego	semáforo	traffic light
(cork)	lente	lens	(bat)	iogurte	yoghurt
	mofo	mold		cereja	cherry
	cone	cone		epiderme	epidermis
borboleta	mangueira	hose	fósforo	cogumelo	mushroom
(butterfly)	obesidade	obesity	(match)	chicote	whip
	eucalipto	eucalyptus		abóbora	pumpkin
	trombone	trombone		alface	lettuce
sanita	geleia	jam	rato	molho	sauce
(toilet)	tarte	pie	(rat)	poste	post
	baloiço	swing		faca	knife
	cabide	hanger		mente	mind
zebra	sarna	scabies	cinzeiro	morango	strawberry
(zebra)	foice	kick	(ashtray)	alfinete	pin
	beiço	lip		bolacha	cookie
	legume	fruit seed		laringe	larynx
chávena	bússola	compass	hipopótamo	candelabro	chandelier
(cup)	sirene	siren	(hippopotamus)	espinafre	spinach
	isqueiro	lighter		cambalhota	back flip
	diamante	diamond		salinidade	salinity
abelha	peruca	wig	prato	circo	circus
(bee)	lactose	propeller	(plate)	iate	yacht
	estojo	case		espada	sword
	quiosque	kiosk		cruz	cross
gravata	cebola	onion	papagaio	antídoto	antidote
(tie)	claque	fan group	(parrot)	sabonete	soap
	centeio	rye		amêndoa	almond
	bosque	grove		gabardine	gabardine
galinha	canela	cinnamon	queijo	sismo	earthquake
(chicken)	semente	seed	(cheese)	baile	ball, dance
	martelo	hammer		relva	grass
	detalhe	detail		elite	elite
carteira	revolta	revolt	caranguejo	termómetro	thermometer
(wallet)	gravidade	gravity	(crab)	piquenique	picnic
	tornado	tornado		trincheira	trench
	recital	recital		amigdalite	tonsillitis
raposa	orelha	ear	casaco	moinho	mill
(fox)	vitrine	glass cabinet	(jacket)	apetite	appetite
	terraço	terrace		barriga	belly
	alicate	cyclone		praxe	hazing
lata	mola	clothespin	dinossauro	telescópio	telescope
(can)	cave	basement	(dinosaur)	chocolate	chocolate
	selo	seal		ferramenta	tool
	dente	tooth		extremidade	limb
girafa	bolota	acorn	aquário	tempero	seasoning
(giraffe)	virose	virosis	(aquarium)	disfarce	disguise
	caroço	pit		pulseira	bracelet

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

abacate	avocado	hélice	propeller

Note. From top to bottom of the distractors columns: gender congruent nouns with the same transparency category of that of the target noun, gender congruent nouns with a different transparency category, gender incongruent nouns with the same transparency category, gender incongruent nouns with a different transparency category.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Table A3

Target and distractor nouns for Experiment 2

Feminine Targets	**Distractors**		**Masculine Targets**	**Distractors**	
bicicleta (bicycle)	camisola	sweater	peixe (fish)	cartaz	exam
	tempestade	storm		evento	event
	catálogo	catalog		frase	phrase
	passaporte	passport		capela	chapel
vaca (cow)	corda	rope	globo (globe)	joelho	knee
	nave	ship		tomate	tomato
	figo	fig		raiva	rage
	lume	fire		raíz	root
toalha (towel)	âncora	anchor	cavalo (horse)	ofício	craft
	couve	cabbage		azeite	olive oil
	granito	granite		placa	board
	cabaz	basket		neve	snow
águia (eagle)	cifra	digit	tapete (rug)	vapor	vapour, steam
	creche	nursery		útero	uterus
	hélio	helium		foice	scythe
	duche	shower		pimenta	pepper
chaminé (chimney)	entorse	sprain	relógio (watch)	tabaco	tobacco
	lágrima	tear		xadrez	chess
	bigode	moustache		manteiga	butter
	teclado	keyboard		humidade	humidity
aranha (spider)	cevada	barley	berço (cradle)	crânio	skull
	estante	bookcase		farol	lighthouse
	gelado	ice cream		tecla	key
	dique	dike		coice	kick
torneira (tap)	almofada	pillow	crocodilo (crocodile)	relâmpago	lightning
	dinamite	dynamite		detergente	detergent
	carimbo	stamp		gabardina	raincoat
	barrete	cap		tangente	tangent
flor (flower)	maré	tide	frigorífico (refrigerator)	contentor	dumpster
	dieta	diet		transplante	transplant
	arroz	rice		recompensa	reward
	ombro	shoulder		tuberculose	tuberculosis
mochila (backpack)	alergia	allergy	armário (closet)	veneno	poison
	cabine	cabin		vinagre	vinegar
	rebanho	flock, herd		lanterna	flashlight
	ciclone	cyclone		diálise	dialysis
tartaruga (turtle)	beterraba	beet	sapato (shoe)	rótulo	tile
	avalanche	avalanche		cacete	baguette
	chouriço	smoked sausage		botija	cylinder
	esparguete	spaghetti		prótese	prosthesis
vassoura (broom)	esmeralda	emerald	copo (glass)	dedo	finger
	cicatriz	scar		poste	post
	intestino	intestine		horta	vegetable garden
	fiambre	ham		peste	plague
avestruz	maionese	mayonnaise	candeeiro	brinquedo	toy

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

(ostrich)	gelatina	jelly	(lamp)	envelope	envelope
	enxame	swarm		basílica	basilica
	remorso	remorse		sanduíche	sandwich
rolha	laca	lacquer	colete	ventre	abdomen
(cork)	lente	lens	(vest)	frasco	bottle
	prado	meadow		lápide	tomstone
	cone	cone		fábula	fable
cadeira	farinha	flour	fósforo	recibo	receipt
(chair)	semente	seed	(match)	chicote	whip
	senado	senate		abóbora	pumpkin
	teste	test		alface	lettuce
sanita	geleia	jam	rato	molho	sauce
(toilet)	tarte	pie	(rat)	pente	comb
	lábio	lip		palha	straw
	cabide	hanger		mente	mind
cama	lenda	legend	cinzeiro	morango	strawberry
(bed)	dor	pain	(ashtray)	revólver	gun
	lago	lake		bolacha	cookie
	gene	gene		laringe	larynx
chávena	bússola	compass	hipopótamo	candelabro	chandelier
(cup)	sirene	siren	(hippopotamus)	espinafre	spinach
	isqueiro	lighter		cambalhota	back flip
	diamante	diamond		salinidade	salinity
cesta	borga	spree (fun)	prato	circo	circus
(basket)	grade	grille	(plate)	iate	yacht
	feno	hay		espada	sword
	vime	wicker		cruz	cross
gravata	cebola	onion	vaso	fungo	fungus
(tie)	claque	fan group	(vase)	pote	bowl
	centeio	rye		telha	roof tile
	bosque	grove		trave	beam
gaveta	prenda	gift	queijo	sismo	earthquake
(drawer)	estirpe	strain	(cheese)	baile	ball, dance
	casino	casino		relva	grass
	arame	wire		elite	elite
carteira	revolta	revolt	termo	navio	ship
(wallet)	gravidade	gravity	(thermos)	exame	exam
	tornado	tornado		loja	store
	recital	recital		carne	flesh
raposa	orelha	ear	casaco	moinho	mill
(fox)	vitrine	glass cabinet	(jacket)	apetite	appetite
	estojo	case		barriga	belly
	alicate	pliers		praxe	hazing
lata	mola	clothespin	dinossauro	telescópio	telescope
(can)	cave	basement	(dinosaur)	chocolate	chocolate
	selo	seal		ferramenta	tool
	dente	tooth		extremidade	limb
garrafa	fogueira	fire pit	aquário	tempero	seasoning
(bottle)	validade	expiration	(aquarium)	vector	vector
	terraço	terrace		pulseira	bracelet

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

tanque	tank	hélice	propeller

Note. From top to bottom of the distractors columns: gender congruent nouns with the same transparency category of that of the target noun, gender congruent nouns with a different transparency category, gender incongruent nouns with the same transparency category, gender incongruent nouns with a different transparency category.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

APPENDIX B: Experimental materials for Study 2

Table B1

Target picture nouns per category in Experiments 1 and 2

Transparent Feminine nouns		**Transparent Masculine nouns**		**Opaque Feminine nouns**		**Opaque Masculine nouns**	
IPNP name	**European Portuguese Translation**	**IPNP name**	**European Portuguese Translation**	**IPNP name**	**European Portuguese Translation**	**IPNP name**	**European Portuguese Translation**
obj005anchor	âncora	obj011arm	braço	obj062bridge	ponte	obj106comb	pente
obj033bathtub	banheira	obj048boat	barco	obj084tape	cassete	obj138drum	tambor
obj053bottle	garrafa	obj050bone	osso	obj094chimney	chaminé	obj204helmet	capacete
obj066butter	manteiga	obj081car	carro	obj097city	cidade	obj261microphone	microfone
obj073camera	câmara	obj085castle	castelo	obj101cloud	nuvem	obj283nose	nariz
obj095church	igreja	obj090cheese	queijo	obj116cross	cruz	obj293paintbrush	pincel
obj104dime	moeda	obj096cigarette	cigarro	obj163flower	flor	obj311pencil	lápis
obj212hose	mangueira	obj180glass	copo	obj225key	chave	obj355ring	anel
obj237leg	perna	obj182globe	globo	obj241lettuce	alface	obj367rug	tapete
obj258mask	máscara	obj203helicopter	helicóptero	obj282net	rede	obj377saxophone	saxofone
obj295palmtree	palmeira	obj281nest	ninho	obj341pyramid	pirâmide	obj431sun	sol
obj339pumpkin	abóbora	obj327plate	prato	obj375sandwich	sanduíche	obj439tank	tanque
obj368ruler	régua	obj352refrigerator	frigorífico	obj418spoon	colher	obj459tomato	tomate
obj392shirt	camisa	obj427strawberry	morango	obj448tennisracket	raquete	obj465tractor	trator
obj491wallet	carteira	obj467train	comboio	obj469tree	árvore	obj484vest	colete
obj510window	janela	obj496watch	relógio	obj492walnut	noz	obj506whip	chicote

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Table B2

Distractors per target picture in each experimental condition in Experiments 1 and 2

Transparent Feminine nouns			**Transparent Masculine nouns**			**Opaque Feminine nouns**			**Opaque Masculine nouns**		
Target	**Distractor**	**(transl.)**	**Target**	**Distractor**	**(transl.)**	**Target**	**Distractor**	**(transl.)**	**Target**	**Distractor**	**(transl.)**
âncora	panela	pan	braço	praia	beach	ponte	feira	fair/market	pente	alga	seaweed
	lábio	lip		turno	shift		vinho	wine		cílio	eyelash
	laje	slab		novidade	novelty		fome	hunger		cárie	caries
	ringue	ring		romance	romance		arte	art		vime	wicker
banheira	compota	jam	barco	praça	square	cassete	esfera	sphere	tambor	pérola	pearl
	amianto	asbestos		dedo	finger		gasóleo	diesel oil		sarampo	measles
	estirpe	strain		tese	thesis		génese	genesis		vaidade	vanity
	trombone	trombone		parque	park		folclore	folklore		rodapé	footer
garrafa	novela	novella	osso	jóia	jewel	chaminé	asneira	swear word	capacete	paródia	parody
	cadeado	padlock		bingo	bingo		donativo	donation		alarido	fuss
	caridade	charity		lente	lens		aeronave	aircraft		avalanche	avalanche
	iene	yen		cume	peak		vinagre	vinegar		frenesim	frenzy
manteiga	reitoria	rectorate	carro	linha	line	cidade	semana	week	microfone	nicotina	nicotine
	estuário	estuary		golo	goal		período	period		labirinto	maze
	humidade	humidity		voz	voice		noite	night		imunidade	immunity
	suspense	thriller		som	sound		gabinete	cabinet		champanhe	champaca
câmara	zona	zone	castelo	avenida	avenue	nuvem	dieta	diet	nariz	órbita	orbit
	efeito	effect		pelouro	office		fígado	liver		alívio	relief
	frente	front		amizade	friendship		ênfase	emphasis		celulose	cellulose
	regime	regime		impasse	deadlock		milagre	miracle		sotaque	accent
igreja	língua	tongue	queijo	parcela	parcel	cruz	fúria	fury	pincel	tigela	bowl
	diálogo	dialogue		prólogo	prologue		lenço	scarf		biombo	screen
	sorte	luck		rapidez	speed		nave	ship		acidez	acidity
	crime	crime		enclave	enclave		iate	yacht		bigode	moustache
moeda	pintura	paint	cigarro	muralha	wall	flor	máfia	mafia	lápis	cebola	onion
	palco	stage		pátio	courtyard		duelo	duel		débito	debit
	greve	strike		interface	interface		maré	tide		cabine	cabin
	convite	invitation		apetite	appetite		arroz	rice		duche	shower
mangueira	cebolada	onion sauce	copo	lema	motto	chave	escola	school	anel	erva	herb
	estrondo	crash		pacto	pact		título	title		boneco	puppet
	anuidade	annuity		peste	plague		ordem	order		foice	scythe
	croquete	croquette		lume	fire		índice	index		boné	cap

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

perna	amora	blackberry	globo	lagoa	pond	alface	pancada	blow	tapete	pimenta	chili
	dardo	dart		ofício	profession		esófago	esophagus		joelho	knee
	raiz	root		índole	character		cirrose	cirrhosis		praxe	initiation ritual
	gene	gene		armazém	warehouse		naipe	suit (cards)		xadrez	chess
máscara	academia	academy	helicóptero	horta	vegetable garden	rede	cabeça	head	saxofone	cafeína	caffeine
	contágio	infection		santuário	sanctuary		preço	price		armário	wardrobe
	barbárie	barbarism		humildade	humility		atitude	attitude		entorse	sprain
	charme	charm		bronze	bronze		nome	name		balancé	swing
palmeira	carroça	wagon	ninho	rampa	ramp	pirâmide	galáxia	galaxy	sol	mesa	table
	ginásio	gym		mastro	mast		charuto	cigar		corpo	body
	grafite	graphite		claque	fan group		metrópole	metropolis		posse	possession
	foguete	rocket		cofre	safe box		chocolate	chocolate		golpe	hit
abóbora	gorjeta	tip	prato	selva	jungle	sanduíche	caderneta	notebook	tanque	agulha	needle
	placebo	placebo		cabelo	hair		vestíbulo	hall		casino	casino
	neurose	neurosis		saudade	yearning		aldrabice	fraud		estante	shelf
	fetiche	fetish		exame	exam		raspanete	telling-off		arame	wire
régua	geleia	jelly	frigorífico	anedota	joke	colher	magia	magic	tomate	comarca	county
	frasco	jar		convénio	pact/agreement		sismo	earthquake		feitiço	spell
	couve	cabbage		planície	lowlands		elite	elite		bondade	goodness
	truque	trick		envelope	envelope		latim	latin		mármore	marble
camisa	relva	grass	morango	trincheira	trench	raquete	alheira	Portuguese sausage	trator	bengala	walking stick
	defeito	defect		estojo	case		espargo	asparagus		fósforo	match
	diocese	diocese		sirene	siren		vigarice	swindle		diálise	dialysis
	ventre	belly		vértice	vertex		filete	fillet		caixote	crate
carteira	placa	board	comboio	fórmula	formula	árvore	lousa	chalkboard	colete	orelha	ear
	oceano	ocean		universo	universe		tornado	tornado		reflexo	reflection
	higiene	hygiene		entidade	entity		tear	loom		lealdade	loyalty
	impacte	impact		jardim	garden		bazar	bazaar		alaúde	lute
janela	estátua	statue	relógio	piscina	pool	noz	família	family	chicote	amêndoa	almond
	navio	ship		cérebro	brain		outubro	october		hálito	breath
	síntese	synthesis		gravidez	pregnancy		imagem	image		marquise	marquee
	pacote	package		bilhete	ticket		futebol	soccer		berlinde	marble (toy)

Note. From top to bottom of the distractors columns: transparent feminine nouns, transparent masculine nouns, opaque feminine nouns, and opaque masculine nouns.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Table B3

Experimental target picture nouns per category in Experiment 3

Opaque Feminine nouns		**Opaque Masculine nouns**	
IPNP name	**European Portuguese Translation**	**IPNP name**	**European Portuguese Translation**
obj062bridge	ponte	obj017baby	bebé
obj084tape	cassete	obj106comb	pente
obj094chimney	chaminé	obj138drum	tambor
obj097city	cidade	obj145elephant	elefante
obj101cloud	nuvem	obj199hat	chapéu
obj116cross	cruz	obj204helmet	capacete
obj163flower	flor	obj226king	rei
obj241lettuce	alface	obj256man	homem
obj282net	rede	obj261microphone	microfone
obj341pyramid	pirâmide	obj283nose	nariz
obj375sandwich	sanduíche	obj338priest	padre
obj418spoon	colher	obj408snail	caracol
obj448tennisracket	raquete	obj431sun	sol
obj469tree	árvore	obj459tomato	tomate
obj492walnut	noz	obj475turkey	peru
obj515woman	mulher	obj157fish	peixe

Table B4

Experimental distractors per target picture in each experimental condition in Experiment 3

Feminine target nouns			**Masculine Target nouns**		
Target	**Distractor**	**Transl.**	**Target**	**Distractor**	**Transl.**
ponte	feira	fair	bebé	letra	letter
	vinho	wine		eixo	axis
	fome	hunger		raiz	root
	sangue	blood		lazer	recreation
cassete	terapia	therapy	pente	lapa	cave/limp
	ginásio	gym		talo	stalk
	génese	genesis		vide	grapevine
	penálti	penalty		rim	kidney
chaminé	asneira	swear word	tambor	gaivota	seagull
	aquário	aquarium		autismo	autism
	aeronave	aircraft		osmose	osmosis
	vinagre	vinegar		clip	clip
cidade	semana	week	elefante	inércia	inertia
	exemplo	example		centeio	rye
	noite	night		lealdade	loyalty
	nível	level		tamboril	monkfish
nuvem	dieta	diet	chapéu	fortuna	fortune
	figado	liver		oceano	ocean

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

	ênfase	emphasis		diocese	diocese
	farol	lighthouse		aluguer	rental
cruz	fúria	fury	capacete	paródia	parody
	quilo	kilo		alarido	fuss
	nave	ship		avalanche	avalanche
	réu	defendant		frenesim	frenzy
flor	máfia	mafia	rei	praia	beach
	duelo	duel		estilo	style
	maré	tide		frase	phrase
	edil	edile		som	sound
alface	gaiola	cage	homem	câmara	camera
	umbigo	belly button		mundo	world
	mascote	pet		crise	crisis
	delfim	dolphin		clube	club
rede	banda	band	microfone	astronomia	astronomy
	preço	price		labirinto	maze
	atitude	attitude		imunidade	immunity
	mar	sea		calcanhar	heel
pirâmide	galáxia	galaxy	nariz	tarefa	assignment
	charuto	cigar		metro	subway
	metrópole	metropolis		défice	deficit
	chocolate	chocolate		convite	invitation
sanduíche	caderneta	notebook	padre	tarefa	assignment
	vestíbulo	hall		metro	subway
	aldrabice	fraud		défice	deficit
	expositor	exhibitor		convite	invitation
colher	magia	magic	caracol	alcova	shelter
	sismo	earthquake		granizo	hail
	elite	elite		benesse	bless
	recital	recital		abdómen	abdomen
raquete	alheira	type of sausage	sol	mesa	table
	canário	canary		corpo	body
	necrose	necrosis		arte	art
	coiote	coyote		casal	couple
árvore	norma	norm	tomate	estátua	statue
	estágio	internship		ozono	ozone
	fraude	fraud		garagem	garage
	cartaz	placard		fervor	fervour
noz	pinça	tweezers	peru	úlcera	ulcer
	veado	deer		cílio	eyelash
	variz	varicose		cárie	caries
	bazar	bazaar		funil	funnel
mulher	família	family	peixe	teoría	theory
	outubro	October		inverno	winter
	imagem	image		lente	lens
	futebol	soccer		café	coffee

Note. From top to bottom of the distractors columns: transparent feminine nouns, transparent masculine nouns, opaque feminine nouns, and opaque masculine nouns.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

APPENDIX C: Experimental materials for Study 3

Table C1

All target nouns according to their translation type

	Target nouns		
	Portuguese	**German**	**English Translation**
Heterogeneric F – M	abóbora	Kürbis	pumpkin
	alface	Salat	salad
	âncora	Anker	anchor
	ambulância	Krankenwagen	ambulance
	árvore	Baum	tree
	barba	Bart	beard
	barriga	Bauch	belly
	cabeça	Kopf	head
	cadeira	Stuhl	chair
	carta	Brief	letter
	cera	Wachs	wax
	colina	Hügel	hill
	dor	Schmerz	pain
	fama	Ruhm	fame
	ferrugem	Rost	rust
	fome	Hunger	hunger
	impressora	Drucker	printer
	lama	Schlamm	mud
	lareira	Kamin	chimney
	lua	Mond	moon
	luva	Handschuh	glove
	maçã	Apfel	apple
	mesa	Tisch	table
	migalha	Krümel	crumb
	pimenta	Pfeffer	pepper
	praia	Strand	beach
	sombra	Schatten	the shade
	tempestade	Sturm	storm
	torneira	Wasserhahn	water tap
	vassoura	Besen	broom
Heterogeneric M – F	moinho	Mühle	mill
	espantalho	Vogelscheuche	scarecrow
	ar	Luft	air
	punho	Faust	fist
	candeeiro	Lampe	lamp
	carvalho	Eiche	oak
	casaco	Jacke	jacket
	chocolate	Schokolade	chocolate
	cigarro	Zigarette	cigarette
	fígado	Leber	liver

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

	figo	Feige	cowardly
	gorro	Mütze	cap
	espectáculo	Veranstaltung	event
	hemisfério	Hemisphäre	hemisphere
	jornal	Zeitung	newspaper
	lábio	Lippe	lip
	leite	Milch	milk
	nabo	Rübe	turnip
	nariz	Nase	nose
	ombro	Schulter	shoulder
	palco	Bühne	stage
	pepino	Gurke	cucumber
	grupo	Gruppe	group
	berço	Wiege	cradle
	relógio	Uhr	clock
	sino	Glocke	bell jar
	sol	Sonne	sun
	teclado	Tastatur	keyboard
	tecto	Decke	ceiling
	tempo	Zeit	time
Homogeneric F	agulha	Nadel	needle
	alergia	Allergie	allergy
	amêndoa	Mandel	almond
	arte	Kunst	arts
	azeitona	Olive	olive
	batata	Kartoffel	potato
	blusa	Bluse	blouse
	farmácia	Apotheke	pharmacy
	cebola	Zwiebel	onion
	cenoura	Karotte	carrot
	chávena	Tasse	cup
	cor	Farbe	colour
	costela	Rippe	rib
	ervilha	Erbse	pea
	escola	Schule	school
	flor	Blume	flower
	fralda	Windel	diaper
	garrafa	Flasche	bottle
	gaveta	Schublade	drawer
	igreja	Kirche	church
	lágrima	Träne	tear
	moeda	Münze	coin
	muleta	Krücke	crutch
	natureza	Natur	nature
	noite	Nacht	night
	noz	Nuss	nut
	parede	Wand	wall

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

	pêra	Birne	pear
	pestana	Wimper	eyelash
	sopa	Suppe	soup
Homogeneric M	açúcar	Zucker	sugar
	alho	Knoblauch	garlic
	ângulo	Winkel	angle
	armário	Schrank	cabinet
	arroz	Reis	rice
	autocarro	Bus	bus
	bosque	Wald	forest
	botão	Knopf	stud
	braço	Arm	poor
	cabo	Kabel	cable
	cachecol	Schal	scarf
	calendário	Kalender	calendar
	cinzeiro	Aschenbecher	ashtray
	círculo	Kreis	circle
	comboio	Zug	train
	cotovelo	Ellbogen	elbow
	espírito	Geist	ghost
	estômago	Magen	stomach
	feitiço	Zauber	magic
	forno	Ofen	oven
	hálito	Atem	breath
	jardim	Garten	garden
	milho	Mais	corn
	musgo	Moos	moss
	osso	Knochen	bone
	queijo	Käse	cheese
	senado	Senat	senate
	tornozelo	Knöchel	ankle
	trigo	Weizen	wheat
	vinagre	Essig	vinegar
Neuter – F	água	Wasser	water
	poema	Gedicht	poem
	bagagem	Gepäck	baggage
	bicicleta	Fahrrad	bicycle
	cama	Bett	bed
	camisa	Hemd	shirt
	cerveja	Bier	beer
	nave	Raumschiff	spaceship
	corda	Seil	rope
	cotonete	Wattestäbchen	cotton swab
	espada	Schwert	sword
	estante	Regal	shelf
	faca	Messer	knife
	farinha	Mehl	flour

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

	ferradura	Hufeisen	horseshoe
	ferramenta	Werkzeug	tool
	fogueira	Lagerfeuer	campfire
	gorjeta	Trinkgeld	tip
	hormona	Hormon	hormone
	idade	Alter	age
	janela	Fenster	window
	lei	Gesetz	law
	loja	Geschäft	shop
	narina	Nasenloch	nostril
	orelha	Ohr	ear
	perna	Bein	leg
	pulseira	Armband	bracelet
	régua	Lineal	ruler
	toalha	Handtuch	hand towel
	trovoada	Gewitter	thunderstorm
Neuter – M	quiosque	Kiosk	kiosk
	ano	Jahr	year
	aquário	Aquarium	aquarium
	átomo	Atom	atom
	crime	Verbrechen	crime
	brinquedo	Spielzeug	toy
	buraco	Loch	hole
	cabelo	Haar	hair
	cacifo	Schließfach	locker
	cano	Rohr	pipe
	carro	Auto	automobile
	cérebro	Gehirn	brain
	copo	Glas	glass
	dinheiro	Geld	money
	esconderijo	Versteck	hiding place
	escritório	Büro	office
	feno	Heu	hay
	fogo	Feuer	fire
	fósforo	Streichholz	match
	golo	Tor	gate
	hospital	Krankenhaus	hospital
	isqueiro	Feuerzeug	lighter
	joelho	Knie	knee
	rumor	Gerücht	rumor
	papel	Papier	paper
	pêndulo	Pendel	pendulum
	terramoto	Erdbeben	earthquake
	teatro	Theater	theatre
	telhado	Dach	roof
	xadrez	Schach	chess

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

APPENDIX D: Information on published papers (*relación de publicacións*)

Ana Rita Sá Leite, Karlos Luna, Isabel Fraga, & Montserrat Comesaña (2020). The gender congruency effect across languages in bilinguals: A meta-analysis. *Psychonomic Bulletin & Review*, *27*(4), 677-693. https://doi.org/10.3758/s13423-019-01702-w (Q1 - 6th Psychology, Experimental – Journal Citation Reports)

Ana Rita Sa Leite conceived the idea and experimental design of the study, conducted the bibliographic search, contacted the authors of the reviewed papers, prepared the data for analysis, was responsible for writing and editing the manuscript overall, and contributed to the main theoretical hypotheses and interpretations. Karlos Luna analysed the data and wrote the result section of the paper. Montserrat Comesaña and Isabel Fraga reviewed the different drafts of the manuscript. All authors made theoretical contributions for the general discussion and approved the final version of the manuscript for submission.

Ana Rita Sá Leite, Juan Haro, Montserrat Comesaña, & Isabel Fraga (2021). Of beavers and tables: The role of animacy in the processing of grammatical gender within a picture-word interference task. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.66117 (Q2 – 43th Psychology, Multidisciplinary – Journal Citation Reports)

Ana Rita Sa Leite and Montserrat Comesaña conceived the idea and experimental design of the study. Ana Rita Sa Leite managed the materials selection as well as the data collection, was responsible for writing and editing the manuscript overall, and contributed to the main theoretical hypotheses and interpretations. Juan Haro analysed the data and wrote the results sections of the paper. Montserrat Comesaña and Isabel Fraga reviewed the different drafts of the manuscript. Isabel Fraga was responsible for the funding of the experiment and its publication. All authors made theoretical contributions for the general discussion and approved the final version of the manuscript for submission.

Ana Rita Sá Leite, Karlos Luna, Ângela Tomaz, Isabel Fraga, & Montserrat Comesaña. (2021*). The underlying mechanisms of grammatical gender selection in language production: A meta-analysis of the gender congruency effect* [Manuscript submitted for publication]. School of Psychology, University of Santiago de Compostela.

Ana Rita Sa Leite conceived the idea and experimental design of the study, conducted the bibliographic search, contacted the authors of the reviewed papers was responsible for writing and editing the manuscript overall, and contributed to the main theoretical hypotheses and interpretations. Ângela Tomaz prepared the data for analysis. Karlos Luna analysed the data and wrote the result section of the paper. Montserrat Comesaña and Isabel Fraga reviewed the different drafts of the manuscript. All authors made theoretical contributions for the general discussion and approved the final version of the manuscript for submission.

Ana Rita Sá Leite, Ângela Tomaz, Juan Andrés Hernández-Cabrera, Isabel Fraga, & Montserrat Comesaña. *What a transparent Romance language with a Germanic gender-determiner mapping tells us about gender retrieval: Insights from European Portuguese* [Manuscript submitted for publication]. School of Psychology, University of Santiago de Compostela.

Ana Rita Sa Leite and Montserrat Comesaña conceived the idea and experimental design of the study. Ana Rita Sa Leite managed the materials selection, was responsible for writing and editing the manuscript overall, and contributed to the main theoretical hypotheses and interpretations. Ângela Tomaz managed data collection. Juan Hernández-Cabrera analysed the data and wrote the results sections of the paper. Montserrat Comesaña and Isabel Fraga reviewed the different drafts of the manuscript. Isabel Fraga and Montserrat Comesaña were responsible for the funding of the experiment. All authors made theoretical contributions for the general discussion and approved the final version of the manuscript for submission.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

APPENDIX E: Permission to use and reproduce published articles

Sá-Leite, Luna, Fraga, & Comesaña (2020). The gender congruency effect across languages in bilinguals: A meta-analysis. *Psychonomic Bulletin & Review*, *27*(4), 677-693. https://doi.org/10.3758/s13423-019-01702-w

SPRINGER NATURE LICENSE
TERMS AND CONDITIONS

Oct 27, 2021

This Agreement between Mrs. Ana Rita Sa Leite Dias ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	5177111349971
License date	Oct 27, 2021
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Psychonomic Bulletin & Review
Licensed Content Title	The gender congruency effect across languages in bilinguals: A meta-analysis
Licensed Content Author	Ana Rita Sá-Leite et al
Licensed Content Date	Jan 21, 2020
Type of Use	Thesis/Dissertation
Requestor type	non-commercial (non-profit)
Format	print and electronic
Portion	full article/chapter
Will you be translating?	no
Circulation/distribution	1 - 29

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Author of this Springer Nature content	yes
Title	REPRESENTATION AND PROCESSING OF GRAMMATICAL GENDER: ANALYSING THE GENDER CONGRUENCY EFFECT
Institution name	Universidade de Santiago de Compostela
Expected presentation date	Nov 2021
Order reference number	1234592a
Requestor Location	Mrs. Ana Rita Sa Leite Dias Calle Xosé María Suárez Núñez, s/n Santiago de Compostela, A Coruña 15782 Spain Attn: Mrs. Ana Rita Sa Leite Dias
Total	0.00 EUR

Terms and Conditions

Springer Nature Customer Service Centre GmbH
Terms and Conditions

This agreement sets out the terms and conditions of the licence (the **Licence**) between you and **Springer Nature Customer Service Centre GmbH** (the **Licensor**). By clicking 'accept' and completing the transaction for the material (**Licensed Material**), you also confirm your acceptance of these terms and conditions.

1. Grant of License

1. 1. The Licensor grants you a personal, non-exclusive, non-transferable, world-wide licence to reproduce the Licensed Material for the purpose specified in your order only. Licences are granted for the specific use requested in the order and for no other use, subject to the conditions below.

1. 2. The Licensor warrants that it has, to the best of its knowledge, the rights to license reuse of the Licensed Material. However, you should ensure that the material you are requesting is original to the Licensor and does not carry the copyright of another entity (as credited in the published version).

1. 3. If the credit line on any part of the material you have requested indicates that it was reprinted or adapted with permission from another source, then you should also seek permission from that source to reuse the material.

2. Scope of Licence

2. 1. You may only use the Licensed Content in the manner and to the extent permitted by these Ts&Cs and any applicable laws.

2. 2. A separate licence may be required for any additional use of the Licensed Material, e.g. where a licence has been purchased for print only use, separate permission must be obtained for electronic re-use. Similarly, a licence is only valid in the language selected and does not apply for editions in other languages unless additional translation rights have been granted separately in the licence. Any content owned by third parties are expressly excluded from the licence.

2. 3. Similarly, rights for additional components such as custom editions and derivatives require additional permission and may be subject to an additional fee. Please apply to Journalpermissions@springernature.com/bookpermissions@springernature.com for these rights.

2. 4. Where permission has been granted **free of charge** for material in print, permission may also be granted for any electronic version of that work, provided that the material is incidental to your work as a whole and that the electronic version is essentially equivalent to, or substitutes for, the print version.

2. 5. An alternative scope of licence may apply to signatories of the STM Permissions Guidelines, as amended from time to time.

3. Duration of Licence

3. 1. A licence for is valid from the date of purchase ('Licence Date') at the end of the relevant period in the below table:

Scope of Licence	Duration of Licence
Post on a website	12 months
Presentations	12 months
Books and journals	Lifetime of the edition in the language purchased

4. Acknowledgement

4. 1. The Licensor's permission must be acknowledged next to the Licenced Material in print. In electronic form, this acknowledgement must be visible at the same time as the figures/tables/illustrations or abstract, and must be hyperlinked to the journal/book's homepage. Our required acknowledgement format is in the Appendix below.

5. Restrictions on use

5. 1. Use of the Licensed Material may be permitted for incidental promotional use and minor editing privileges e.g. minor adaptations of single figures, changes of format, colour and/or style where the adaptation is credited as set out in Appendix 1 below. Any

other changes including but not limited to, cropping, adapting, omitting material that affect the meaning, intention or moral rights of the author are strictly prohibited.

5. 2. You must not use any Licensed Material as part of any design or trademark.

5. 3. Licensed Material may be used in Open Access Publications (OAP) before publication by Springer Nature, but any Licensed Material must be removed from OAP sites prior to final publication.

6. Ownership of Rights

6. 1. Licensed Material remains the property of either Licensor or the relevant third party and any rights not explicitly granted herein are expressly reserved.

7. Warranty

IN NO EVENT SHALL LICENSOR BE LIABLE TO YOU OR ANY OTHER PARTY OR ANY OTHER PERSON OR FOR ANY SPECIAL, CONSEQUENTIAL, INCIDENTAL OR INDIRECT DAMAGES, HOWEVER CAUSED, ARISING OUT OF OR IN CONNECTION WITH THE DOWNLOADING, VIEWING OR USE OF THE MATERIALS REGARDLESS OF THE FORM OF ACTION, WHETHER FOR BREACH OF CONTRACT, BREACH OF WARRANTY, TORT, NEGLIGENCE, INFRINGEMENT OR OTHERWISE (INCLUDING, WITHOUT LIMITATION, DAMAGES BASED ON LOSS OF PROFITS, DATA, FILES, USE, BUSINESS OPPORTUNITY OR CLAIMS OF THIRD PARTIES), AND
WHETHER OR NOT THE PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. THIS LIMITATION SHALL APPLY NOTWITHSTANDING ANY FAILURE OF ESSENTIAL PURPOSE OF ANY LIMITED REMEDY PROVIDED HEREIN.

8. Limitations

8. 1. ***BOOKS ONLY:*** Where **'reuse in a dissertation/thesis'** has been selected the following terms apply: Print rights of the final author's accepted manuscript (for clarity, NOT the published version) for up to 100 copies, electronic rights for use only on a personal website or institutional repository as defined by the Sherpa guideline (www.sherpa.ac.uk/romeo/).

8. 2. For content reuse requests that qualify for permission under the STM Permissions Guidelines, which may be updated from time to time, the STM Permissions Guidelines supersede the terms and conditions contained in this licence.

9. Termination and Cancellation

9. 1. Licences will expire after the period shown in Clause 3 (above).

9. 2. Licensee reserves the right to terminate the Licence in the event that payment is not received in full or if there has been a breach of this agreement by you.

USC DE SANTIAGO DE COMPOSTELA

Appendix 1 — Acknowledgements:

For Journal Content:
Reprinted by permission from [**the Licensor**]: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication)

For **Advance Online Publication papers:**
Reprinted by permission from [**the Licensor**]: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication), advance online publication, day month year (doi: 10.1038/sj.[JOURNAL ACRONYM].)

For Adaptations/Translations:
Adapted/Translated by permission from [**the Licensor**]: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication)

Note: For any republication from the British Journal of Cancer, the following credit line style applies:

Reprinted/adapted/translated by permission from [**the Licensor**]: on behalf of Cancer Research UK: : [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication)

For **Advance Online Publication** papers:
Reprinted by permission from The [**the Licensor**]: on behalf of Cancer Research UK: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication), advance online publication, day month year (doi: 10.1038/sj.[JOURNAL ACRONYM])

For Book content:
Reprinted/adapted by permission from [**the Licensor**]: [**Book Publisher** (e.g. Palgrave Macmillan, Springer etc) [**Book Title**] by [**Book author**(s)] [**COPYRIGHT**] (year of publication)

Other Conditions:

Version 1.3

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.

Sá-Leite, Haro, Comesaña, & Fraga (2021). Of beavers and tables: The role of animacy in the processing of grammatical gender within a picture-word interference task. *Frontiers in Psychology, 12*. https://doi.org/10.3389/fpsyg.2021.66117

Frontiers allows the re-use and adaptation of their published papers for commercial or non-commercial purposes, given that attribution to the original article published in Frontiers is made explicit in the new work. Please, check: https://www.frontiersin.org/about/policies-and-publication-ethics.

"1. Frontiers Policies

1.1. Open Access and Copyright

All Frontiers articles from July 2012 onwards are published with open access under the CC-BY Creative Commons attribution license (the current version is CC-BY, version 4.0). This means that the author(s) retains copyright, but the content is free to download, distribute, and adapt for commercial or non-commercial purposes, given appropriate attribution to the original article.

Upon submission, the author(s) grants Frontiers a license to publish, including to display, store, copy, and reuse the content. The CC-BY Creative Commons attribution license enables anyone to use the publication freely, given appropriate attribution to the author(s) and citing Frontiers as the original publisher. The CC-BY Creative Commons attribution license does not apply to third-party materials that display a copyright notice to prohibit copying. Unless the third-party content is also subject to a CC-BY Creative Commons attribution license, or an equally permissive license, the author(s) must comply with any third-party copyright notices."

.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Resumo en galego

O xénero gramatical é unha das características máis desconcertantes da linguaxe tanto dende a perspectiva da Lingüística como da Psicolingüística. O seu estudo levou a multitude de autores a facer variadas afirmacións nos que reflicten a súa frustración en relación a el en termos pexorativos. Así por exemplo, o xénero foi descrito como superficial, prescindible e carente de significado, así como de propósito (Fodor, 1959; Ibrahim, 1973; Kilarski, 2005). Quizais destaca o moi gráfico símil de Trudgill ao referirse ao xénero gramatical como "pezóns masculinos lingüísticos" ("*linguistic male niples*", 1999; p. 149), así como a desesperanzadora conclusión de Boutonnet et al. (2012) de que o xénero simplemente escapa á lóxica. En efecto, aínda que a función e funcionamento desta característica na linguaxe non están esclarecidos o certo é que o xénero gramatical existe en aproximadamente a metade das linguas do mundo (Corbett, 2013) e está presente de maneira determinante na maioría dos niveis de organización e codificación destas linguas. De feito, unha vez é adquirido, o noso sistema lingüístico non pode escapar del, o que levou a moitos psicolingüistas a interesarse pola ligazón entre xénero gramatical e cognición. Se os filósofos nun inicio e os lingüistas despois xa entraran en multitude de discusións co xénero coma protagonista, os psicolingüistas non quedaron atrás: co auxe desta disciplina xurdiron acaloradas discusións sobre a representación e procesamento do xénero na mente. En particular, mentres que a adquisición do xénero gramatical e o seu papel na comprensión lingüística son temas xa minimamente clarificados a día de hoxe, o procesamento do xénero durante a produción lingüística conforma unha área repleta de resultados contraditorios e teorías confrontadas, un verdadeiro caos científico que é o foco de interese da presente tese. A grandes trazos, as disputas céntranse especialmente na suposta obrigatoriedade dos contextos de concordancia de xénero para a recuperación do valor de xénero dun substantivo, na natureza automática ou competitiva dos mecanismos detrás da selección de xénero, e nos diferentes puntos de vista sobre a ligazón entre a morfo-fonoloxía nominal e o xénero.

USC DE COMPOSTELA

Para dar resposta a estas preguntas, abordamos a literatura sobre o xénero gramatical utilizando unha perspectiva integral. Tras unha breve introdución na cal explicamos a estrutura da tese, no capítulo 2 definimos que é o xénero dende a lingüística, destacando os dous polos de asignación e concordancia, marcando a diferenza cos clasificadores, e rematando coa distinción máis propia da psicolingüística entre xénero natural e gramatical, para despois centrar o foco de atención do lector neste último. No capítulo 3 abordamos o concepto de transparencia de xénero con base na literatura sobre a adquisición da linguaxe e sobre a comprensión lingüística. Acabamos propoñendo, por unha banda, unha definición clara de que é transparencia de xénero baseada na fonoloxía e morfoloxía de xénero e na coincidencia entre ambas, e pola outra, un método estandarizado que pretende operacionalizar o grao de transparencia de xénero das linguas a través dunha adaptación do concepto de complexidade relativa dos sistemas de xénero de Audring (2019). Desta maneira, aspiramos a que todas as linguas poidan ser avaliadas seguindo dito método e situadas nun continuo de transparencia como o proposto por Küpisch et al. (2018). No capítulo 4 expoñemos os dous modelos clásicos de produción da linguaxe que incluíron o xénero gramatical nas súas propostas: o modelo Word Encoding by Activation and VERification (WEAVER++, Levelt et al., 1999) e o modelo *Independent Network* (IN, Caramazza, 1997). Os dous entenden que o acceso ao léxico durante a produción de substantivos dáse a través da codificación conceptual nun nivel semántico, a codificación gramatical e sintáctica nun nivel sintáctico ou lema, e a codificación orto-fonolóxica nun nivel da forma ou lexema. Ambos representan o xénero gramatical como nodos de xénero (un nodo por cada valor) ao cal os substantivos están conectados dependendo do seu valor de xénero. Estes nodos atópanse no nivel sintáctico ou lema. De acordo co modelo WEAVER++, o procesamento sintáctico (e do xénero) precede ao da forma, aínda que pode haber certo solapamento por procesamento en cascada. Importante, este modelo concibe a selección de xénero como un proceso de competición entre os dous nodos, pero distingue "activación" de "selección": os nodos de xénero actívanse cando procesamos un substantivo, pero soamente cando o xénero é necesario para a concordancia (para determinar a forma de palabras como artigos ou adxectivos) esta activación acada o limiar de selección e por tanto, o xénero é recuperado. O modelo IN entende que o procesamento no nivel sintáctico soamente ocorre despois do procesamento da representación orto-fonolóxica do nome. A selección do xénero gramatical sería un proceso competitivo, no cal se inhibiría o nodo non-seleccionado. Este modelo non se pronuncia acerca do papel dos contextos de concordancia.

USC UNIVERSIDADE DE SANTIAGO

No capítulo 5 abordamos o núcleo da tese: a literatura experimental sobre a recuperación e procesamento do xénero gramatical durante o acceso a substantivos en produción lingüística no marco dos dous modelos anteriores. Para abordar a temática, consideramos dúas perspectivas: a produción de línguas nativas vs. segundas linguas (L2s).

En canto ao estudo do procesamento do xénero gramatical durante a produción de linguas nativas, este realizouse de maneira pioneira a través do xa clásico paradigma de interferencia imaxe-palabra. Nesta tarefa, os participantes deben nomear imaxes en voz alta utilizando un nome illado ("casa") ou unha frase nominal ("a casa") ao mesmo tempo que ignoran un substantivo escrito sobre a imaxe. O substantivo *target* (o que representa a imaxe) e o substantivo *distractor* (o que deben ignorar) poden ter o mesmo ou distinto valor de xénero. Os investigadores esperaban detectar variacións nos tempos de nomeado dos participantes dependendo desa congruencia de xénero, en concreto, esperaban que os *targets* acompañados de *distractors* do mesmo xénero desen lugar a tempos de resposta menores (efecto de congruencia de xénero, ECX). Isto porque ambos estarían activando o mesmo nodo de xénero na nosa mente, facilitando a chegada ao limiar de activación necesario para a selección de dito nodo. O certo é que a cuestión distou de ser tan simple como parecía. No caso dos falantes nativos de linguas xermánicas e eslavas, o ECX soamente é observado cando os participantes nomean as imaxes utilizando frases nominais formadas por nomes e outras palabras cuxa forma varía dependendo do xénero do nome, isto é, elementos de concordancia de xénero (p.ex., Bordag & Pechmann, 2008; Bürki et al., 2016; Heim, Friederici, et al., 2009; Schiller & Caramazza, 2003; Schiller & Costa, 2006). Pola contra, cando os participantes usan nomes illados ou frases nominais cuxo elemento que acompaña ao nome non ten que concordar en xénero (en concreto, no alemán e no holandés o artigo definido en plural non varía de acordo co xénero), non se observa ningún tipo de efecto competitivo (p.ex., Finocchiaro et al., 2011; La Heij et al, 1998; Schiller & Caramazza, 2006; Starreveld & La Heij, 2004). Isto orixinou dúas grandes interpretacións confrontadas: unha que entende o ECX como competición entre os nodos de xénero activados por *target* e distractor (p.ex., feminino vs. masculino) e outra que entende o ECX como competición entre os posibles determinantes a ser seleccionados (p.ex., no alemán, "*der*" [artigo definido masculino no singular] vs. "*die*" [artigo definido feminino]). A primeira, apóiase no modelo WEAVER++, que di que se a concordancia de xénero non é necesaria, a nosa mente non recupera o valor de xénero dun nome, e de aí a ausencia de efectos neste paradigma cando non hai elementos que varíen dun xénero a outro. A segunda apóiase na

USC UNIVERSIDADE DE SANTIAGO

evidencia obtida polo chamado "paradigma singular-plural", que mostra que mesmo no plural o xénero está sendo recuperado (o que desmonta a premisa do WEAVER++) e, ademais, leva as súas afirmacións cara o extremo: segundo os autores que defenden esta hipótese, como a selección dos valores de xénero non produce efectos xenuínos de competición, a súa recuperación é automática e non está suxeita a competición (Caramazza et al., 2001).

No caso da literatura con linguas romances, o escenario é aínda menos claro. Por unha banda, cando os participantes utilizan frases nominais (con elementos de concordancia de xénero) para nomear as imaxes, non se observan efectos de competición de ningún tipo (p.ex., Costa et al., 1999; Cubelli et al., 2005; Miozzo & Caramazza, 1999; Miozzo et al., 2002). Por outra banda, cando utilizan nomes illados, obsérvanse ou ben un efecto en dirección contraria en italiano e castelán (substantivos co mesmo xénero dificultan a resposta – incrementan os tempos de nomeado; efecto de incongruencia de xénero, EIX; Cubelli et al., 2005; Paolieri et al., 2010, 2011), ou ben efectos nulos nesas mesmas linguas e tamén en francés (Finocchiaro et al. 2011; O'Rourke, 2007). En relación á ausencia de resultados con frases nominais, a literatura focalízase na hipótese de selección tardía (HST). De acordo con esta, dada a complexidade do paradigma de interferencia imaxe-palabra, que se apoia na interacción entre procesos de recoñecemento e produción para estudar a codificación de características en produción, este non sería capaz de captar o proceso de competición entre *target* e *distractor* cando hai un proceso de competición posterior que enmascara ese outro proceso máis temperán. Concretamente, os autores explican que en linguas como o italiano e o castelán existe máis dunha posible forma de artigo definido por cada valor de xénero (en castelán, "*la*" e "*el*" para o feminino dependendo da acentuación e forma do nome; en italiano, "*lo*", "*le*", "*l'*" para o masculino dependendo das primeiras letras da seguinte palabra, que pode ser ou ben un nome ou un adxectivo). Por esta razón, para que un artigo sexa seleccionado, o sistema ten que procesar antes información adicional que en linguas cun mapeado de xénero-determinante un-a-un non é necesaria (como no alemán, no cal o sistema soamente necesita procesar "definido", "singular", e "masculino" para seleccionar ao "*der*"). Estes procesos adicionais de procesamento da palabra seguinte e de identificación e organización dos elementos na oración enmascaran o proceso de competición que ocorreu nun inicio, e nesa etapa tardía as que compiten en todo caso son as formas dos artigos para un mesmo xénero, e non as formas dos artigos activadas por distintas palabras (*target* e *distractor*). En efecto, en francés tampouco se obtivo o ECX en condicións clásicas; sen embargo, nesta lingua o mapeado xénero-forma é un

USC UNIVERSIDADE DE SANTIAGO

tanto distinto: existe un artigo definido por valor de xénero, e unha forma reducida igual para ambos xéneros no caso dos substantivos que comezan por vogal ("*le*" e "*la*" como "*l'*"). Foucart et al. (2010) indican que no caso do francés, como a transformación non depende do xénero en si, soamente son necesarios procesos de transformación posteriores que retrasarían a selección do artigo pero non no mesmo grao que en linguas como o italiano e o castelán. Por iso, argumentan que a presentación máis tardía do *distractor* en relación ao *target* podería resolver este enmascaramento, como así demostraron evidenciando o ECX con SOAs (*stimulus onset asynchrony*) negativos. Bürki et al. (2020) non logran replicar este resultado, pero si observan o ECX cando teñen en conta nos modelos de análise finais os tempos de lectura dos distractores e non só os de resposta.

En canto á situación observada con nomes illados, a interpretación dos resultados é aínda mais arrevesada. Por un lado, é difícil encadrar o chamado EIX dentro da estrutura do acceso ao léxico proposta por modelos clásicos de produción lingüística e apoiada por variada evidencia empírica. De acordo co modelo de dobre ruta, os que están competindo pola selección son os nomes en si mesmos, e non os valores de xénero, isto porque o efecto mostra que pares de nomes do mesmo xénero dificultan a resposta do participante. Co obxectivo de explicar os devanditos resultados, Cubelli et al. (2005) desenvolven un modelo, o modelo de dupla selección, que se apoia en dúas premisas: (1) o xénero gramatical non fai parte dun nivel de procesamento gramatical e sintáctico (o lema) separado do nivel conceptual, se non que o primeiro nivel de procesamento de acceso ao léxico é semántico-sintáctico, de maneira que o xénero é unha característica semántica máis; (2) existe unha descomposición morfo-fonolóxica pola cal a raíz dunha palabra é procesada a través da información semántica do significado e o final da palabra a través da característica semántica do xénero (p.ex., en "*casa*": "*cas-*" e "*-a*", sendo que "*-a*" se procesa porque "*casa*" é feminina). Ademais, Cubelli et al. (2005) propoñen que o EIX é observado en italiano e castelán e non en linguas xermánicas e eslavas debido a que as primeiras teñen unha alta proporción de nomes transparentes baseados nas pistas "-a" e "-o" e as outras non. Certamente, o modelo de dupla selección é importante no sentido de que é o primeiro en considerar as peculiaridades de cada lingua ou familia lingüística, introducindo o concepto de transparencia (fonolóxica) de xénero no debate sobre produción lingüística. Sen embargo, foi amplamente criticado por igualar fonoloxía de xénero a morfoloxía de xénero (dicir que é igual o "*-a*" de "*cas**a***" e o de "*chic**a***"); deixar sen explicar claramente que ocorre coas palabras opacas e irregulares (con outras terminacións non asociadas claramente a un valor

USC UNIVERSIDADE DE SANTIAGO

de xénero, como "-e", e con terminacións propias dun xénero sendo que son de outro valor de xénero, como o substantivo "*mano*" [man], que é feminino), e propoñer unha estrutura semántico-sintáctica que non termina de encaixar coa evidencia existente (ver Finocchiaro et al., 2011).

En definitiva, os resultados con falantes de linguas romances son altamente controvertidos, porque (1) no caso das frases nominais, poñen aparentemente de manifesto a incapacidade do paradigma de interferencia imaxe-palabra para detectar certos procesos de competición (HST); (2) no caso dos nomes illados, os resultados contradinse e, particularmente no que se refire ao EIX, este vai contra a evidencia anterior doutras áreas e contra a noción da estrutura de acceso ao léxico; (3) non permiten apoiar claramente unha hipótese de competición (valores de xénero *vs*. determinantes) sobre a outra.

Co fin de dilucidar os efectos baseados no paradigma de interferencia imaxe-palabra, realizamos unha meta-análise na que incluímos 39 experimentos de 14 estudos diferentes e calculamos o tamaño do efecto de xénero (a diferenza entre as condicións de xénero congruente e incongruente, expresada xa sexa como "ECX" ou "EIX") a través da *g* de Hedges. Analizamos, ademais, a posible presenza dun nesgo de publicación. Tivemos en consideración as variables familia lingüística (romance vs. xermánica ou eslava), transparencia de xénero da lingua (transparentes vs. opacas), presenza de concordancia de xénero (nome illado vs. frase nominal con elemento de concordancia) e tipo de concordancia de xénero (elementos de concordancia libres vs. morfemas). Os resultados poñen de manifesto que: (1) nas linguas xermánicas e eslavas, mentres que o "ECX" entendido como competición entre determinantes (elementos libres) é de tamaño medio e significativo ($g = 0.424$; $p < .001$), o ECX sen contexto de concordancia non é significativo; (2) o ECX soamente é significativo con elementos libres de concordancia de xénero, pero non cando a concordancia é feita a través de morfemas; (3) nas linguas romance, o EIX é extremadamente pequeno ($g = -0.171$) pero significativo ($p = .021$); e (4) non hai nesgo de publicación nin no ECX entendido como competición entre determinantes, nin no EIX. Realizamos unha segunda meta-análise para comprobar, a través dun paradigma máis simple -o paradigma do singular-plural- a existencia de competición entre morfemas de xénero. Isto permitiría, por unha banda, confirmar os mecanismos de selección destes elementos e, por outra , comparar o tamaño do efecto deste paradigma máis simple có do paradigma de interferencia imaxe-palabra. Así dispoñeríamos de maior información sobre o posible enmascaramento de efectos que a complexidade deste paradigma de interferencia pode

USC UNIVERSIDADE DE SANTIAGO

carrear. Sen embargo, o número de estudos existentes non nos permitiu obter resultados que puidesen acadar a significación, o que supón un obstáculo para poder sacar conclusións en relación a estes temas.

En resumo, a literatura co paradigma de interferencia imaxe-palabra suxire que os "ECX" que se poden obter son de competición entre determinantes, mentres que non está claro o motivo de observar o EIX con nomes illados nalgunhas linguas romance. Ademais, hai sospeitas de que o paradigma de interferencia imaxe-palabra constitúe unha tarefa complexa que pode non ser o suficientemente sensible para detectar certos efectos (véxase o que ocorre en italiano e castelán segundo a HST). É importante destacar aquí que Heim, Eickhoff, et al. (2009) e Heim, Friederici, et al. (2009) empregaron técnicas de medición máis sensibles (fMRI) con nativos de alemán e observaron modificacións na resposta hemodinámica asociada ás condicións de congruencia de xénero que suxerían, pola área á que estaba asociada, competición entre valores de xénero, e non entre determinantes, quizais mostrando unha fase máis temperá da competición *target – distractor*.

A literatura con poboación bilingüe, pola súa parte, baséase en tarefas de nomeado de imaxes simple (p.ex., dicir "Baum" en alemán [L2, en castelán -L1- "*árbol*"] ante a imaxe dunha árbore) ou en tarefas de tradución (p.ex., dicir "*Baum*" en alemán [L2] ante a palabra "*árbol*" en castelán [L1]). As condicións de congruencia e incongruencia de xénero son por tanto determinadas pola coincidencia do valor de xénero entre traducións. Na grande maioría dos estudos, un ECX interlingüístico é atopado de maneira consistente independentemente da familia lingüística e do tipo de instrución de nomeado (nome illado *vs*. frase nominal con elementos de concordancia; p.ex., Bordag & Pechmann, 2007; Klassen, 2016; Lemhöfer et al., 2008; Manolescu & Jarema, 2015; Morales et al., 2016; Paolieri et al., 2019). En consecuencia, os resultados apoian claramente, por unha parte, unha concepción compartida do sistema de xénero das linguas do bilingüe (isto é, as entradas léxicas están conectadas aos mesmos nodos independentemente da lingua, o que xera efectos competitivos de selección) e, por outra, a hipótese de que os efectos de competición baséanse nos valores de xénero, e que, por tanto, os nodos de xénero compiten entre si pola selección, en contra das ideas de Caramazza et al. (2001).

USC UNIVERSIDADE DE SANTIAGO

Para establecer as conclusións da literatura en produción bilingüe, realizamos unha meta-análise na que incluímos 22 experimentos de 10 artigos e calculamos o tamaño do ECX interlingüístico a través da *g* de Hedges. Medimos o impacto das variables familia

lingüística/transparencia de xénero da lingua (opaca vs. transparente), tipo de instrución de nomeado ou tradución (nome illado vs. frase nominal), competencia na L2, e tipo de tarefa (nomeado de imaxes vs. tradución). Os resultados poñen de manifesto que o ECX interlingüístico é pequeno e variable (g = .24; $Q_E(34)$ = 190.27) pero significativo (p < .001). Por tanto, apoian a idea de que o sistema de xénero do bilingüe é compartido entre linguas e que os nodos de xénero compiten entre si, sendo que esta competición non desaparece en función da lingua producida ou da concordancia de xénero. Tampouco as variables competencia na L2 e tipo de tarefa influíron nos resultados, se ben os resultados no caso da primeira deben collerse con precaución: a literatura xeral en bilingüismo dinos que a competencia debería ter afectado os resultados. De feito, unha inspección dos estudos incluídos na meta-análise mostra que o modo de medición de dito parámetro en cada estudo varía amplamente, sendo moitas veces subxectivo.

No Capítulo 6 da tese introducimos no debate unha lingua que non fora considerada previamente na literatura: o portugués, en particular, o portugués europeo. Así, antes de establecer conclusións sobre a extensa revisión teórica e meta-analítica presentada e tratar de dar resposta ás cuestións formuladas no inicio, levamos a cabo unha serie de experimentos co paradigma de interferencia imaxe-palabra con falantes nativos de portugués, así como un experimento de tradución bidireccional considerando bilingües temperáns de portugués e alemán. O portugués é unha lingua interesante de cara ao debate xa que se trata dunha lingua romance e transparente, similar ao italiano e ao castelán, pero cun mapeado de xénero-determinantes un-a-un en relación aos artigos definidos, de maneira similar ás linguas xermánicas. Isto permite, por un lado, poñer a proba a HST, xa que se a ausencia de ECX con frases nominais en linguas romances se debe a algunha outra diferenza coa familias de linguas xermánicas e eslavas, o efecto tampouco debería atoparse en portugués. Por outro, permite testar a replicabilidade do EIX, xa que os preceptos do modelo de dupla selección deberían aplicarse ao portugués tal e como se aplican ao castelán e ao italiano, porque é tamén unha lingua coas pistas de xénero "*-o*" e "*-a*". Ademais, propuxémonos: (1) aumentar o control experimental na selección de materiais en relación á literatura anterior; (2) controlar debidamente e estudar directamente dúas variables usualmente ignoradas: a transparencia orto-fonolóxica de xénero (cunha adaptación propia do modelo de dobre ruta de Gollan e Frost [2001] á produción lingüística), e a animacidade; (3) medir de maneira estandarizada con tests de alta fiabilidade e validez a competencia nas linguas dos nosos participantes bilingües.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Así, despois de realizar unha análise profunda do sistema de xénero do portugués europeo, mostrando o seu alto grao de transparencia de xénero, expoñemos os estudos experimentais da tese: o Estudo 1, composto por dous experimentos de interferencia imaxe-palabra no que pedimos aos participantes que nomearan as imaxes usando nomes illados e manipulamos a cantidade de *targets* animados presentes na lista de estímulos (0%, 25%, 50%, 100%); o Estudo 2, composto por tres experimentos de interferencia imaxe-palabra nos que pedimos aos participantes que nomearan as imaxes usando frases nominais (Experimento 1) e nomes illados (Experimentos 2 e 3), introducindo no deseño non só a congruencia de xénero senón tamén a transparencia orto-fonolóxica de xénero de *targets* e distractores; o Estudo 3, composto por un experimento de tradución bidireccional cunha mostra de bilingües temperáns de portugués e alemán onde se mide rigorosamente e inclúese nos modelos finais a competencia tanto nunha lingua como na outra a través do Dialang PT e Dialang Ger. Resumidamente, as análises realizadas baseáronse en *linear mixed models* e os resultados mostraron: (1) un ECX no Estudo 1 na condición con 0% de *targets* animados (Experimento 1) que pon de relevo o papel da inclusión de estímulos animados en experimentos deste tipo por razóns de priorización semántica e nesgo atencional - este ECX sen embargo, non é confirmado polas análises Bayesianas realizadas e non foi replicado nos Experimentos 2 e 3 do Estudo 2; (2) un ECX entendido como competición entre determinantes no Experimento 1 con frases nominais do Estudo 2, apoiando a HST - este ECX, sen embargo, é mais complexo có encontrado en linguas xermánicas e eslavas, xa que interactúa coa transparencia de xénero, sendo que os *targets* opacos son os principais responsables do efecto, especialmente cando van emparellados con distractores transparentes. Isto apoia a idea de que as rotas de procesamento de xénero atopadas en comprensión lingüística tamén operan en produción lingüística; (3) un EXC interlingüístico con nomes illados entendido como interferencia entre traducións de diferente valor de xénero que aumenta conforme se incrementa a discrepancia entre os graos de competencia das dúas linguas – o que apoia en última instancia, un sistema de xénero compartido entre linguas, a competición entre valores de xénero durante o proceso de selección do xénero, e o papel determinante da competencia das linguas nesta competición.

Tendo en consideración a revisión da literatura realizada, as meta-análises presentadas, e o noso propio traballo experimental co portugués europeo, chegamos ás seguintes conclusións: (A) os efectos condutuais obtidos no paradigma de interferencia imaxe-palabra encaixan coa hipótese da competición entre determinantes, e non entre valores de xénero; (B) isto non

significa que o xénero non estea competindo: cabe aos investigadores evitar interpretar a ausencia de evidencia como evidencia da ausencia dun efecto (Vasishth & Gerlman, 2021); (C) o EIX continúa sendo problemático en termos de interpretación, pero, en calquera caso, cando xuntamos os nosos propios estudos á meta-análise, o EIX pasa a mostrar unha significación marxinal, polo que é necesaria a replicación futura de estudos similares para determinar a robustez de dito efecto; (D) os nosos resultados amosan certa variabilidade inter-individual, xa que a dirección dos potenciais efectos parece variar de participante a participante; (E) a transparencia de xénero parece desempeñar un papel na selección dos valores de xénero en produción lingüística; (F) o mapeado xénero-determinantes un-a-varios proposto pola HST está detrás da ausencia de efectos con linguas como o italiano ou o castelán; (G) os efectos observados coas tarefas bilingües reflicten unha competición xenuína entre valores de xénero que depende da competencia en ambas linguas.

Este conxunto de resultados suxire que (1) o paradigma máis empregado de estudo das cuestións de xénero gramatical en produción lingüística (paradigma de interferencia imaxe-palabra) quizá non sexa o máis adecuado; (2) por esta razón, sospeitamos que os posibles efectos de xénero – que son en todo caso pequenos e difíciles de detectar, tal e como mostra a meta-análise bilingüe- poden estar enmascarados polo paradigma, quedando desenmascarados en estudos como o de Heim, Friederici, et al. (2009); (3) aínda así, esta é unha hipótese que debe acumular evidencia na investigación futura e debe ser contrastada coa hipótese de non-competición dos valores de xénero; (4) para poder, polo tanto, abordar ambas hipóteses de maneira adecuada, urxe a replicación de estudos coas mesmas e con novas linguas, o coidado de aspectos como o tamaño das mostras e o control dos materiais, a consideración de variables como a transparencia de xénero, o uso de técnicas de medida máis precisas, como a fMRI ou a electroencefalografía, e quizais, incluso, un cambio de paradigma; (5) é moi probable que non haxa unha resposta única ás preguntas que nos formulamos no inicio, xa que a natureza da representación e o procesamento gramatical do xénero tende a estar dirixida por características específicas das linguas, do número de linguas faladas e outras variables de natureza psicolingüística.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Referencias:

Audring, J. (2019). Canonical, complex, complicated? En F. Di Garbo, B. Olsson & B. Wälchli (Eds.), *Grammatical gender and linguistic complexity: Volume I: General issues and*

specific studies (Studies in Diversity Linguistics 26) (pp. 15–52). Language Science Press.

Bordag, D., & Pechmann, T. (2007). Factors influencing L2 gender processing. *Bilingualism: Language and Cognition*, *10*, 299–314. https://doi.org/10.1017/s1366728907003082

Bordag, D., & Pechmann, T. (2008). Grammatical gender in speech production: Evidence from Czech. *Journal of Psycholinguistic Research*, *37*(2), 69–85. https://doi.org/10.1007/s10936-007-9060-0

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

Bürki, A., Besana, T., Degiorgi, G., Gilbert, R., & Alario, F. -X. (2020). Representation and selection of determiners with phonological variants. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(7), 1287–1315. https://doi.org/10.1037/xlm0000643

Bürki, A., Sadat, J., Dubarry, A.-S., & Alario, F. -X. (2016). Sequential processing during noun phrase production. *Cognition*, *146*, 90–99. https://doi.org/10.1016/j.cognition.2015.09.002

Caramazza, A., & Miozzo, M. (1997). The relation between syntactic and phonological knowledge in lexical access: Evidence from the "tip-of-the-tongue" phenomenon. *Cognition*, *64*(3), 309–343. https://doi.org/10.1016/S0010-0277(97)00031-0

Caramazza, A., Miozzo, M., Costa, A., Schiller, N. O., & Alario, F.-X. (2001). A Cross-linguistic Investigation of Determiner Production. En E. Dupoux (Ed.), *Language, brain, and cognitive development: Essays in honor of Jacques Mehler* (pp. 209–226). MIT Press.

Corbett. G. G. (2013). Number of Genders. En M. S. Dryer & M. Haspelmath (Eds.), *The World Atlas of Language Structures Online*. Max Planck Institute for Evolutionary Anthropology. Recuperado de http://wals.info/chapter/30

Costa, A., Sebastián-Gallés, N., Miozzo, M., & Caramazza, A. (1999). The gender congruity effect: Evidence from Spanish and Catalan. *Language and Cognitive Processes*, *14*(4), 381–391. https://doi.org/10.1080/016909699386275

Cubelli, R., Lotto, L., Paolieri, D., Girelli, M., & Job, R. (2005). Grammatical gender is selected in bare noun production: Evidence from the picture–word interference paradigm.

USC UNIVERSIDADE DE SANTIAGO

Journal of Memory and Language, *53*(1), 42–59. https://doi.org/10.1016/j.jml.2005.02.007

Finocchiaro, C., Alario, F. X., Schiller, N. O., Costa, A., Miozzo, M., & Caramazza, A. (2011). Gender congruency goes Europe: A cross-linguistic study of the gender congruency effect in Romance and Germanic languages. *Rivista di Linguistica*, *23*(2), 161–198.

Fodor, I. (1959). The origin of grammatical gender. *Lingua*, *8*, 186–214. https://doi.org/10.1016/0024-3841(59)90020-8

Foucart, A., Branigan, H. P., & Bard, E. G. (2010). Determiner selection in romance languages: Evidence from French. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(6), 1414–1421. https://doi.org/10.1037/a0020432

Heim, S., Eickhoff, S. B., Friederici, A. D., & Amunts, K. (2009). Left cytoarchitectonic area 44 supports selection in the mental lexicon during language production. *Brain Structure and Function*, *213*(4-5), 441–456. https://doi.org/10.1007/s00429-009-0213-9

Heim, S., Friederici, A. D., Schiller, N. O., Rüschemeyer, S. A., & Amunts, K. (2009). The determiner congruency effect in language production investigated with functional MRI. *Human brain mapping*, *30*(3), 928–940. https://doi.org/10.1002/hbm.20556

Ibrahim, M. H. (1973). *Grammatical Gender: Its Origin and Development*. De Gruyter Mouton. https://doi.org/10.1515/9783110905397

Kilarski, M. (2005). On grammatical gender as an arbitrary and redundant category. En D. A. Kibbee (Ed.), *History of Linguistics 2005: Selected papers from the Tenth International Conference on the History of the Language Sciences (ICHOLS X), 1–5 September 2005, Urbana-Champaign, Illinois* (pp. 24–36). John Benjamins Publishing Company. https://doi.org/10.1075/sihols.112.03kil

Klassen, R. (2016). The representation of asymmetric grammatical gender systems in the bilingual mental lexicon. *Probus*, *28*(1). https://doi.org/10.1515/probus-2016-0002

Kupisch, T., Geiß, M., Mitrofanova, N., & Westergaard, M. (2018*). Gender Cues in L1 Russian Children Acquiring German as an Early L2*. EuroSLA 28 - Universität Münster. Münster, Alemaña.

Lemhöfer, K., Schriefers, H., & Jescheniak, J. D. (2006). The processing of free and bound gender-marked morphemes in speech production: Evidence from Dutch. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(2), 437–442. https://doi.org/10.1037/0278-7393.32.2.437

USC UNIVERSIDADE DE SANTIAGO

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. https://doi.org/10.1017/S0140525X99001776

Manolescu, A., & Jarema, G. (2015). Grammatical gender in Romanian-French bilinguals. *The Mental Lexicon*, *10*, 390–412. https://doi.org/10.1075/ml.10.3.04man

Miozzo, M., & Caramazza, A. (1999). The selection of determiners in noun phrase production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(4), 907–922. https://doi.org/10.1016/S0749-596X(02)00531-4

Miozzo, M., Costa, A., & Caramazza, A. (2002). The absence of a gender congruency effect in romance languages: A matter of stimulus onset asynchrony? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(2), 388–391. https://doi.org/10.1037/0278-7393.28.2.388

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: Language and Cognition*, *19*(2), 294–310. https://doi.org/10.1017/s1366728915000176

O'Rourke, P. L. (2007). The gender congruency effect in bare noun production in Spanish. *Coyote Papers: Working Papers in Linguistics, Linguistic Theory at the University of Arizona*, *15*, 66–89.

Paolieri, D., Lotto, L., Leoncini, D., Cubelli, R., & Job, R. (2011). Differential effects of grammatical gender and gender inflection in bare noun production. *The British Psychological Society*, *102*(1), 19–36. https://doi.org/10.1342/000712610X496536

Paolieri, D., Lotto, L., Morales, L., Bajo, M. T., Cubelli, R., & Job, R. (2010). Grammatical gender processing in romance languages: Evidence from bare noun production in Italian and Spanish. *European Journal of Cognitive Psychology*, *22*(3), 335–347. https://doi.org/10.1080/09541440902916803

Paolieri, D., Padilla, F., Koreneva, O., Morales, L., & Macizo, P. (2019). Gender congruency effects in Russian–Spanish and Italian–Spanish bilinguals: The role of language proximity and concreteness of words. *Bilingualism: Language and Cognition*, *22*, 112–129. https://doi.org/10.1017/s1366728917000591

Schiller, N. O., & Caramazza, A. (2003). Grammatical feature selection in noun phrase production: Evidence from German and Dutch. *Journal of Memory and Language*, *48*(1), 169–194. https://doi.org/10.1016/S0749-596X(02)00508-9

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Starreveld, P., & La Heij, W. (2004). Phonological facilitation of grammatical gender retrieval. *Language and Cognitive Processes*, *19*(6), 677–711. https://doi.org/10.1080/01690960444000061

Trudgill, P. (1999). Language contact and the function of linguistic gender. *Poznań Studies in Contemporary Linguistics*, *35*, 133–152. Recuperado de http://ifa.amu.edu.pl/psicl/files/35/10Trudgill.pdf

Vasishth, S., & Gelman, A. (2021). How to embrace variation and accept uncertainty in linguistic and psycholinguistic data analysis. *Linguistics*. https://doi.org/10.31234/osf.io/zcf8s

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA